STABILITY ANALYSIS OF HETEROGENEOUS HELMHOLTZ PROBLEMS AND FINITE ELEMENT SOLUTION BASED ON PROPAGATION MEDIA APPROXIMATION

HÉLÈNE BARUCQ, THÉOPHILE CHAUMONT-FRELET, AND CHRISTIAN GOUT

Abstract. The numerical simulation of time-harmonic waves in heterogeneous media is a tricky task which consists in reproducing oscillations. These oscillations become stronger as the frequency is increasing and high-order finite element methods have demonstrated their capability to reproduce the oscillatory behavior. However they keep coping with limitations in capturing fine scale heterogeneities. We propose a new approach which can be applied in highly heterogeneous propagation media. It consists in constructing an approximate medium in which we can perform computations for a large variety of frequencies. The construction of the approximate medium can be understood as applying a quadrature formula locally. We establish estimates which generalize existing estimates formerly obtained for homogeneous Helmholtz problems. We then provide numerical results which illustrate the good level of accuracy of our solution methodology.

Introduction

Wave propagation is a complex physical phenomenon which is involved in a large number of applications like for instance radar or sonar detection, medical or seismic imaging. Numerical simulations for waves deserve attention because they require applying advanced numerical methods in particular when the propagation domain is heterogeneous. First, there is a need in tracing the wave frequencies accurately and that may be a tricky task because frequencies can be wide-ranging for many applications. For example, depth imaging is searching for deeper layers which may contain hydrocarbons and it uses frequencies which must be of a few tens of Hertz with a very low resolution. If it is to detect hidden objects, the depth of the explored region does not exceed a few tens of meters and the involved frequencies are close to the kilohertz. High performing numerical methods should thus be stable for a widest as possible frequency range. In particular, these methods should minimize phenomena of numerical pollution that generate errors increasing faster with frequency than with the inverse of space discretization step. Next, heterogeneities heavily impact the behavior of waves. Numerical methods must then be able to take them into account. A medium can be heterogeneous in different ways. For instance, it can be stratified and highly contrasted. It can also include very small heterogeneities as compared with the size of the domain. In each case, the characteristics of the propagation medium can be described by the variations of the velocity parameter and in most of the realistic cases, it is not even continuous.

High-order methods have become very popular to discretize wave problems because they let reducing the pollution effect and thus considering high frequencies [12, 22, 23]. They involve high degree polynomial basis functions which are built on coarse meshes. High frequencies can thus be considered but strong heterogeneities are not taken into account correctly. For that reason, high-order methods do not perform as well as possible in highly heterogeneous media. Indeed, if simulations are to have any chance reproducing waves inside heterogeneous media, the size of mesh cells must be small enough to capture the heterogeneities [1].

2000 *Mathematics Subject Classification.* Primary 65N12, 65N15, 65N30; Secondary 35J05.

Key words and phrases. Helmholtz equation, highly heterogeneous media, high order method, medium approximation, stability estimates, depth imaging.

This work was partially supported by the project M2NUM ("Grand Réseaux de Recherche" of the Région Normandie and co-financed by FEDER funds) and the INRIA-TOTAL strategic action DIP (`http://dip.inria.fr`).

The authors also thank the Centre for Computer Resources of Normandy (CRIANN, `http://www.criann.fr`) where numerical simulations have been done (in parts).

When the constitutive parameters are highly oscillating, like in composite materials, homogenization techniques are attractive and we can mention a very recent work not yet published but available on line [14]. It provides a mathematical analysis of the Helmholtz equation with highly oscillating velocity parameter when it is solved with Finite Element Heterogeneous Multiscale Method. In the same spirit, Capdeville et al. [5] have proposed an upscaling tool which is based on a two-scale homogenization expansion and provides a modelling valid in heterogeneous elastic media. It is a non periodic procedure which has been tested with spectral elements and it gives accurate results. Nevertheless, convergence analysis has never been delivered. Furthermore, the homogenized medium is obtained as the solution of an auxiliary problem which might be challenging to solve, even with advanced dedicated techniques [6].

In this paper, we adopt another point of view avoiding homogenization techniques. We propose a subcell approximation strategy which makes it possible to handle very small heterogeneities on a coarse mesh, even if high degree polynomial basis functions are used. We call this approach the Multiscale Medium Approximation method (MMAm).

We restrict our analysis to continuous FEM schemes because it is simpler to present, but it is based on general arguments which can be applied to other mesh-based discretization strategies such as the discontinuous Galerkin methods presented in [2, 3, 9]. Though the analysis of the present paper is limited to the linear case, numerical examples show that higher order polynomial approximations are working well with subcell variation of the velocity. The analysis of higher order polynomial FEM is currently under investigation.

To the best of our knowledge, this paper is the first attempt to discretize the heterogeneous Helmholtz equation with jump in the wavespeed inside the mesh cells. Indeed, even if recent advances have been made in the context of plane wave methods, they are either limited by velocity-fitting meshes [15,18,19], i.e. the wavespeed is constant inside each cell, or to smooth wavespeeds [13,24].

Since we consider a general velocity parameter, the entries of the corresponding finite element linear system can not be computed exactly. The key idea of the MMAm is to consider an approximation c_h of the actual velocity parameter c for which analytical expression of the entries are available. It turns out that approximating the propagation medium amounts to use a quadrature formula particularly designed for the actual velocity. Then, we prove that the MMAm is stable and convergent if the approximation c_h converges to c, where h stands for the local space step approximation. We develop a two-scale (h, H) convergence analysis where H is the finite element mesh step. This is done by extending the recent convergence and stability results based on elliptic projection of Wu and Zhu [25,26].

To carry out the convergence analysis of the MMAm, frequency-explicit stability estimates of the continuous problem are required. It seems that such stability estimates are not available in the literature. Indeed, To the best of our knowledge, even though recent advances have been made, existence, uniqueness and stability results for Helmholtz equation with mixed boundary condition and an arbitrary parameter $c \in L^\infty(\Omega)$ are not available in the literature. For instance, Lechleiter and collaborators successfully show the well-posedness of a heterogeneous Helmholtz and Maxwell problems in [16,17], but the velocity parameter has to be C^0 or $W^{1,\infty}$ regular and the stability constants are not optimal in frequency. Sharper estimates have been obtain in [11], but they require H^2 regularity of the right hand side.

We propose to fill this lack by extending the works presented in [8,10] providing a technical assumption on the velocity parameter c (see condition (2)). This hypothesis is only made for technical reasons but it is representative of lots of geological media corresponding to a local approximation of the Earth as a stratified-like medium. Our analysis is also limited to the case where the domain is surrounded by an absorbing boundary condition, however, numerical experiments show that the numerical method perform as well in more complex media.

The guideline of this paper is the following. Section 1 presents a stability analysis for the Helmholtz problem set in an heterogeneous media. Section 2 then aims at showing that it is possible to take the discontinuities of c into account on a coarse mesh by considering an approximation c_h. We then show at Section 3 that c_h can be chosen to obtain a quadrature-like formula that can be mastered to ensure the construction of the discrete system is cheap. The paper ends up with

Section 4 which is devoted to numerical experiments in two dimensional domains illustrating the results obtained at Sections 1 and 2. In particular, we show that the MMAm outperforms standard finite element approximations in highly heterogeneous media, even when technical assumption (2) is not satisfies.

1. Analytical study

This section deals with a stability analysis of the Helmholtz equation set in a heterogeneous medium that we have chosen to represent by the variations of the velocity for which a technical assumption is required. We deem it technical because we have no evidence that it is mandatory. In particular, numerical experiments present the same behaviour if the hypothesis is not satisfied. Our analysis covers existence, uniqueness, stability and regularity results, in particular at high frequency, and it is carried out as an adaptation of [10] and [21] to heterogeneous media.

Let Ω be the propagation domain. Having in mind regional seismic wave simulations, we assume Ω is a rectangle but our work is easily extendable to other geometries. We thus define $\Omega = (0, L_1) \times (0, L_2) \subset \mathbb{R}^2$, where the vertical axis is oriented from the top to the bottom. The propagation of harmonic seismic waves is governed by the Helmholtz equation:

$$\left\{ \begin{array}{rcll} -k^2 u - \Delta u & = & f & \text{in } \Omega \\ \nabla u \cdot \mathbf{n} - ik_{max} u & = & 0 & \text{on } \partial\Omega. \end{array} \right. \tag{1}$$

The wave number k is defined from the pulsation ω and the velocity c through the relation $k = \omega/c$. The pulsation is a given positive constant and the velocity varies in the whole domain and $k_{max} = \sup_\Omega k$. Since we are especially concerned with high frequency waves, we consider pulsations ω higher than a given minimum ω_0. The field f is a given distributed source. To get into the right condition of numerical experiments, we assume that the domain of interest is limited by an absorbing boundary. We thus set the simplest outgoing radiation condition on the boundary of Ω.

In the case where k is constant, this problem has been analysed and it is well-known that for any f in $L^2(\Omega, \mathbb{C})$, the problem is well-posed in the sense of Hadamard (see [10] for instance). This paper pertains the case where k is variable following the variations of the velocity. We assume that $c \in L^\infty(\Omega)$ is piecewise constant and the values of c are distributed as follows. The velocity model is composed of R subdomains Ω_r enclosed in Ω and in each Ω_r, the velocity is $c_r = c|_{\Omega_r} \in \mathbb{R}^{+*}$ with $c_{min} = \min_r c_r$, $c_{max} = \max_r c_r$ and we assume that $c_{min} > 0$. We further assume that there exists a point $x_0 \in \Omega$ such that

$$\frac{\mathbf{n}_r \cdot (x - x_0)}{c_r^2} + \frac{\mathbf{n}_l \cdot (x - x_0)}{c_l^2} < 0 \quad \forall x \in \Omega_r \cap \Omega_l, \tag{2}$$

for all $r, l \in \{1, \ldots, R\}$ such that $\Omega_r \cap \Omega_l \neq \emptyset$. Examples of velocity models satisfying (2) are given in Figures 1 and 2.

In the following, we employ the notation $k_r = \omega/c_r$. We also adopt standard notations for the functional spaces, norms and inner products and their definitions can be found for instance in Ciarlet [7].

It is well-known that $u \in H^1(\Omega, \mathbb{C})$ is solution to (1) in a weak sense if and only if u satisfies the variational equation

$$B(u, v) = -\int_\Omega k^2 u\bar{v} - ik_{max} \int_{\partial\Omega} u\bar{v} + \int_\Omega \nabla u \cdot \nabla \bar{v} = \int_\Omega f\bar{v}, \tag{3}$$

for all $v \in H^1(\Omega, \mathbb{C})$, where $B : H^1(\Omega, \mathbb{C}) \times H^1(\Omega, \mathbb{C}) \to \mathbb{C}$ is the sesquilinear form associated with Problem (1).

Proposition 1. *Let $u \in H^1(\Omega, \mathbb{C})$ be any solution to (3). Then $u \in H^2(\Omega, \mathbb{C})$ and there exists a constant $C := C(\Omega, c_{min})$ such that*

$$|u|_{2,\Omega}^2 \leq C\big(|f|_{0,\Omega}^2 + (\omega^2 + \omega^4)|u|_{0,\Omega}^2 + \omega^2 |u|_{1,\Omega}^2\big).$$

Proof. u being a solution to (3), it satisfies

$$\int_\Omega \nabla u \cdot \nabla \bar{v} = \int_\Omega F\bar{v} + \int_{\partial\Omega} G\bar{v} \qquad \forall v \in H^1(\Omega, \mathbb{C}),$$

with $F = f + k^2 u$ and $G = ik_{max}u$. Since $k \in L^\infty(\Omega)$ and $u \in H^1(\Omega, \mathbb{C})$, we have $F \in L^2(\Omega, \mathbb{C})$ and $G \in H^{1/2}(\partial\Omega, \mathbb{C})$. Since Ω is convex, the classical theory for the homogeneous Laplace operator implies that there exists a constant C depending on Ω only such that

$$|u|^2_{2,\Omega} \leq C\big(|F|^2_{0,\Omega} + ||G||^2_{1/2,\partial\Omega}\big).$$

Furthermore, regarding norms $|F|^2_{0,\Omega}$ and $||G||^2_{1/2,\partial\Omega}$, we have

$$\begin{aligned} |F|^2_{0,\Omega} &= |f + k^2 u|^2_{0,\Omega} \\ &\leq |f|^2_{0,\Omega} + k^4_{max}|u|^2_{0,\Omega} \\ &\leq C\big(|f|^2_{0,\Omega} + \omega^4 |u|^2_{0,\Omega}\big), \end{aligned}$$

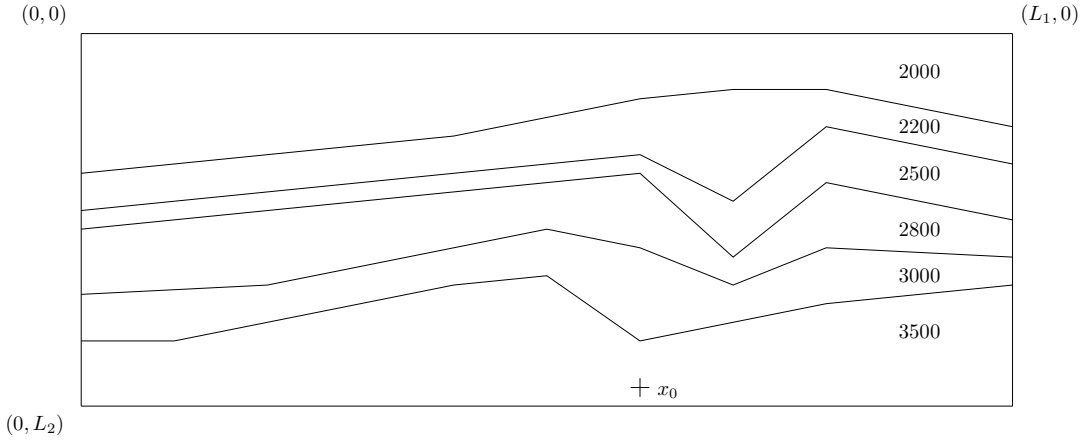


FIGURE 1. A stratified velocity parameter

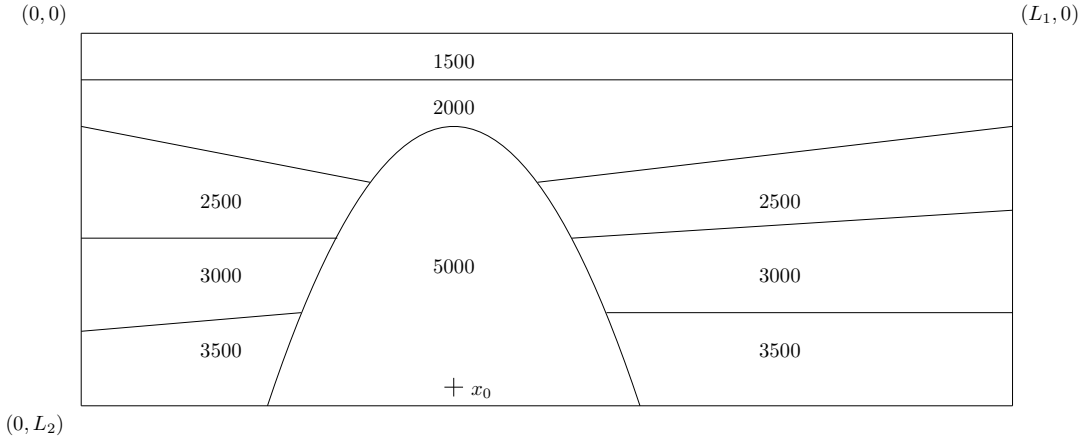


FIGURE 2. A velocity parameter with a salt body

with $C = \max(1, 1/c_{min}^4)$. Moreover,

$$\begin{aligned} ||G||^2_{1/2,\partial\Omega} &= ||ik_{max}u||^2_{1/2,\partial\Omega} \\ &= k^2_{max}||u||^2_{1/2,\partial\Omega} \\ &\leq C\omega^2||u||^2_{1/2,\partial\Omega}, \end{aligned}$$

with $C = \max(1, 1/c_{min}^2)$. We end the proof thanks to the following trace inequality

$$||u||^2_{1/2,\Omega} \leq C\big(|u|^2_{0,\Omega} + |u|^2_{1,\Omega}\big),$$

where C is a constant depending on Ω only. □

Before turning to stability in the $L^2(\Omega, \mathbb{C})$ norm, we state two identities which are established in the Appendix. The first one is the classical Rellich identity: for all $w \in H^2(\Omega, \mathbb{C})$,

$$2\text{Re} \int_\Omega \nabla w \cdot (\mathbf{x} \cdot \nabla \bar{w}) = \int_{\partial\Omega} |\nabla w|^2 \mathbf{x} \cdot \mathbf{n}. \tag{4}$$

The second identity reads as:

Lemma 1. *For all $w \in H^1(\Omega, \mathbb{C})$,*

$$2Re \int_\Omega k^2 w\mathbf{x} \cdot \nabla \bar{w} = -2 \int_\Omega k^2 |w|^2 + \sum_{r,l=1}^{R} \int_{\Omega_r \cap \Omega_l} (k_r^2 \mathbf{x} \cdot n_r + k_l^2 \mathbf{x} \cdot n_l)|w|^2 + \int_{\partial\Omega} k^2 |w|^2 \mathbf{x} \cdot \mathbf{n}, \tag{5}$$

where $\mathbf{x} = x - x_0$.

Proposition 2. *Let $u \in H^1(\Omega, \mathbb{C})$ be any solution to (3). Then there exists a constant $C := C(\Omega, c_{max}, c_{min}, x_0, \omega_0)$ such that*

$$|u|_{0,\Omega} \leq \frac{C}{\omega} |f|_{0,\Omega}.$$

Proof. According to Proposition 1, $u \in H^2(\Omega, \mathbb{C})$ and $v = \mathbf{x} \cdot \nabla u$ is regular enough to be used as a test function in the variational equation (3). Recalling (4) and (5), then taking the real part of (3), we have

$$\begin{aligned} 2 \int_\Omega k^2 |u|^2 - \sum_{r,l=1}^{R} \int_{\Omega_r \cap \Omega_l} (k_r^2 \mathbf{x} \cdot n_r + k_l^2 \mathbf{x} \cdot n_l)|u|^2 + \int_{\partial\Omega} |\nabla u|^2 \mathbf{x} \cdot \mathbf{n} \\ = 2\text{Re} \int_\Omega f \mathbf{x} \cdot \nabla \bar{u} + 2\text{Re}\ ik_{max} \int_{\partial\Omega} u \mathbf{x} \cdot \nabla \bar{u} + \int_{\partial\Omega} k^2 |u|^2 \mathbf{x} \cdot \mathbf{n}. \end{aligned}$$

Ω being a rectangle, it is strictly star-shaped with respect to x_0, and there exists a constant $\gamma > 0$ depending on Ω and x_0 only such that $\mathbf{x} \cdot \mathbf{n} \geq \gamma$ on $\partial\Omega$. Since c satisfies (2), we have $(k_r^2 \mathbf{x} \cdot \mathbf{n}_r + k_l^2 \mathbf{x} \cdot \mathbf{n}_l) \leq 0$. Then, observing that $|\mathbf{x}| \leq \text{diam } \Omega = (L_1^2 + L_2^2)^{1/2} = L$, it follows

$$\begin{aligned} 2k_{min}^2 |u|^2 + \gamma |\nabla u|^2_{0,\partial\Omega} &\leq 2L|f|_{0,\Omega}|u|_{1,\Omega} + 2Lk_{max}|u|_{0,\partial\Omega}|\nabla u|_{0,\partial\Omega} + Lk^2_{max}|u|^2_{0,\partial\Omega} \\ &\leq \frac{L^2}{\epsilon}|f|^2_{0,\Omega} + \epsilon|u|^2_{1,\Omega} + \frac{L^2 k^2_{max}}{\gamma}|u|^2_{0,\partial\Omega} + \gamma|\nabla u|^2_{0,\Omega} + Lk^2_{max}|u|^2_{0,\partial\Omega}. \end{aligned}$$

We then get that for any $\epsilon > 0$

$$2k_{min}|u|^2_{0,\Omega} \leq \frac{L^2}{\epsilon}|f|^2_{0,\Omega} + \epsilon|u|^2_{1,\Omega} + \left(\frac{L^2}{\gamma} + L\right) k^2_{max}|u|^2_{0,\partial\Omega}. \tag{6}$$

We complete the proof by deriving estimates for $|u|_{1,\Omega}$ and $|u|_{0,\partial\Omega}$. This is carried out by picking $v = u$ as a test function in (3) and considering the real and imaginary parts separately. We start by pertaining $|u|_{1,\Omega}$. We have:

$$\text{Re } B(u,u) = -\int_\Omega k^2|u|^2 + \int_\Omega |\nabla u|^2 = \text{Re} \int_\Omega f\bar{u} \leq |f|_{0,\Omega}|u|_{0,\Omega}.$$

It follows that

$$|u|_{1,\Omega}^2 \leq |f|_{0,\Omega}|u|_{0,\Omega} + k_{max}^2|u|_{0,\Omega}^2 \leq \frac{1}{4k_{max}^2}|f|_{0,\Omega}^2 + 2k_{max}^2|u|_{0,\Omega}^2.$$

Then selecting $\epsilon_0 = k_{min}^2/4k_{max}^2$, we obtain

$$\frac{L^2}{\epsilon_0}|f|_{0,\Omega} + \epsilon_0|u|_{1,\Omega}^2 \leq \left(4L^2\frac{k_{max}^2}{k_{min}^2} + \frac{k_{min}^2}{k_{max}^4}\right)|f|_{0,\Omega}^2 + \frac{k_{min}^2}{2}|u|_{0,\Omega}^2. \tag{7}$$

We now move on estimating $|u|_{0,\partial\Omega}$. We have

$$\text{Im } B(u,u) = -k_{max}|u|_{0,\partial\Omega}^2 = \text{Im } \int_\Omega f\bar{u}.$$

It follows that

$$\begin{aligned}\left(\frac{L^2}{\gamma} + L\right)k_{max}^2|u|_{0,\partial\Omega}^2 &\leq \left(\frac{L^2}{\gamma} + L\right)k_{max}|f|_{0,\Omega}|u|_{0,\Omega} \\ &\leq \frac{1}{2}\left(\frac{L^2}{\gamma} + L\right)^2\frac{k_{max}^2}{k_{min}^2}|f|_{0,\Omega}^2 + \frac{k_{min}^2}{2}|u|_{0,\Omega}^2. \end{aligned} \tag{8}$$

Combining (6), (7) with (8), we get

$$k_{min}^2|u|^2 \leq \left\{\left(4L^2 + \frac{1}{2}\left(\frac{L^2}{\gamma} + L\right)^2\right)\frac{k_{max}^2}{k_{min}^2} + \frac{k_{min}^2}{k_{max}^4}\right\}|f|_{0,\Omega}^2,$$

so that the proposition holds with

$$C = c_{max}\sqrt{\left(4L^2 + \frac{1}{2}\left(\frac{L^2}{\gamma} + L\right)^2\right)\frac{c_{min}^2}{c_{max}^2} + \frac{c_{max}^2}{c_{min}^4}\frac{1}{\omega_0^2}}.$$

□

We end this section by a full statement of the results obtained in the section.

Theorem 1. *Problem* (3) *admits a unique solution* $u \in H^1(\Omega,\mathbb{C})$. *Furthermore,* $u \in H^2(\Omega,\mathbb{C})$, *and there exists a constant* $C := C(\Omega, c_{min}, c_{max}, x_0, \omega_0)$ *such that*

$$|u|_{0,\Omega} \leq \frac{C}{\omega}|f|_{0,\Omega}, \quad |u|_{1,\Omega} \leq C|f|_{0,\Omega}, \quad |u|_{2,\Omega} \leq C\omega|f|_{0,\Omega}.$$

Proof. Regarding existence and uniqueness, observe that the sesquilinear form B satisfies a Gårding inequality. Indeed for all $v \in H^1(\Omega,\mathbb{C})$, we have

$$\text{Re } B(v,v) = -\int_\Omega k^2|v|^2 + \int_\Omega |\nabla v|^2 \geq -k_{max}^2|v|_{0,\Omega}^2 + |v|_{1,\Omega}^2.$$

Therefore, it follows that we can apply the Fredholm alternative and thus focus on uniqueness. But Proposition 2 applied to (3) with $f = 0$ implies that $u = 0$, which proves uniqueness and thus existence.

Problem (3) admits thus a unique solution $u \in H^1(\Omega,\mathbb{C})$. Now, Proposition 2 implies that

$$|u|_{0,\Omega} \leq \frac{C_0}{\omega}|f|_{0,\Omega},$$

with a suitable constant C_0. Moreover, we have

$$\text{Re } B(u,u) = -\int_\Omega k^2|u|^2 + \int_\Omega |\nabla u|^2 = \text{Re } \int_\Omega f\bar{u}.$$

which implies that

$$\begin{aligned}
|u|_{1,\Omega}^2 &\leq |f|_{0,\Omega}|u|_{0,\Omega} + k_{max}^2|u|_{0,\Omega}^2 \\
&\leq \frac{1}{4k_{max}^2}|f|_{0,\Omega}^2 + 2k_{max}^2|u|_{0,\Omega}^2 \\
&\leq \left(\frac{c_{min}^2}{4\omega_0^2} + 2\frac{C_0^2}{c_{min}^2}\right)|f|_{0,\Omega}^2 \\
&\leq C_1^2|f|_{0,\Omega}^2.
\end{aligned}$$

The demonstration of the theorem is then ended since Proposition 1 let us write the estimates:

$$\begin{aligned}
|u|_{2,\Omega}^2 &\leq C(\Omega)\big(|f|_{0,\Omega}^2 + (\omega^2+\omega^4)|u|_{0,\Omega}^2 + \omega^2|u|_{1,\Omega}^2\big) \\
&\leq C(\Omega)\big(1 + (1+\omega^2)C_0^2 + \omega^2C_1^2\big)|f|_{0,\Omega}^2 \\
&\leq C(\Omega)\big(\frac{1+C_0^2}{\omega^2} + (C_0^2+C_1^2)\big)\omega^2|f|_{0,\Omega}^2 \\
&\leq C(\Omega)\big(\frac{1+C_0^2}{\omega_0^2} + (C_0^2+C_1^2)\big)\omega^2|f|_{0,\Omega}^2 \\
&\leq C_2^2\omega^2|f|_{0,\Omega}^2.
\end{aligned}$$

□

Corollary 1. *Consider $g \in L^2(\Omega,\mathbb{C})$. Then, there exists a unique element $z \in H^1(\Omega,\mathbb{C})$ satisfying the adjoint equation*

$$B^*(z,w) = \overline{B(w,z)} = \int_\Omega g\bar{w} \quad \forall w \in H^1(\Omega,\mathbb{C}).$$

Furthermore, $z \in H^2(\Omega,\mathbb{C})$ and there exists a constant $C := (\Omega, c_{min}, c_{max}, x_0, \omega_0)$ such that

$$|z|_{0,\Omega} \leq \frac{C}{\omega}|g|_{0,\Omega}, \quad |z|_{1,\Omega} \leq C|g|_{0,\Omega}, \quad |z|_{2,\Omega} \leq C\omega|g|_{0,\Omega}.$$

Proof. The sesquilinear form B is not self-adjoint but B and B^* are closely reading. Indeed, observe that for all $u, v \in H^1(\Omega,\mathbb{C})$, we have

$$B^*(z,w) = \overline{B(w,z)} = -\int_\Omega k^2 w\bar{z} + ik_{max}\int_{\partial\Omega} w\bar{z} + \int_\Omega \nabla w\cdot\nabla\bar{z}.$$

The two sesquilinear forms thus differ by the sign before the boundary integral only, and as a matter of fact, all the demonstrations given for problem (3) apply with minor modifications to the adjoint problem. □

2. Numerical analysis

In this section, we pertain a finite element discretization of problem (3) and we study its convergence with respect to the pulsation ω and the maximum size H of cells forming the mesh. In the case of homogeneous domains, the condition $\omega^2 H < C$ is known to be suboptimal. Many authors have proposed different proofs to obtain sharper stability conditions. For example, by using a numerical Green function to the 1D Helmholtz problem set in a homogeneous domain, Ilhenburg and Babŭska have shown in [12] that the condition $\omega^{2p+1}H^{2p} < C$ is sufficient, where p denotes the degree of polynomial functions that are used for approximation. In higher dimensions, Melenk and Sauter have used a frequency splitting argument to demonstrate in [22, 23] that the condition $\omega^{p+1}H^p < C$ is sufficient even if the datum $f \in L^2(\Omega)$ is rough.

We propose convergence estimates which are based on the analysis of Zhu and Wu [25,26]. Our proof is elaborated for a 2D heterogeneous domain and its main ingredient is the construction of an approximate propagation medium by the mean of an approximate velocity c_h. We are then able to extend the optimal convergence result for linear elements in homogeneous media providing that ω^3H^2 and $\omega\mathcal{M}_{H,h}$ are small enough. The quantity $\mathcal{M}_{H,h}$ which involves two parameters H and h, stands for the approximation error of c by c_h (see definition 2). As abovementionned, H

denotes the discretization step related to the finite element mesh while h represents the size of the local submesh that is used to represent the approximate velocity c_h.

Let then $\mathcal{T}_H$ be a regular mesh of Ω and its associated conforming discrete space $V_H \subset H^1(\Omega, \mathbb{C})$. Since Theorem 1 indicates that $u \in H^2(\Omega, \mathbb{C})$, we may expect a linear convergence in the $H^1(\Omega, \mathbb{C})$ norm when using linear or bilinear elements.

We now tackle the issue of computing the entries of the linear system associated with V_H. Indeed, even when using piecewise polynomials, we must integrate quantities involving c. In fact, if we assume that each interface $\Omega_r \cap \Omega_l$ is polygonal, we could accurately mesh it with a finer mesh $\mathcal{T}_h$ where h has already been introduced with the approximate velocity c_h. We could then perform an exact integration on $\mathcal{T}_h$. But this is not fully satisfactory since it requires to build an auxiliary mesh and we prefer to avoid any superfluous mesh with a view to reduce the implementation time. Furthermore, if we accept the idea of constructing an auxiliary mesh, the quadrature scheme induced by the fine mesh $\mathcal{T}_h$ is different in each coarse cell, making integration of linear system entries very costly. Finally, for realistic applications, the interfaces $\Omega_l \cap \Omega_r$ are not given explicitly and the parameter c is rather given as a set of sampling values. It seems thus difficult to introduce $\mathcal{T}_h$. We have to cope with a technical difficulty and for that purpose, we propose to construct an approximation c_h of c such that the entries of the linear system are both cheap and easy to compute. This is what we are doing in Section 3 but before, we focus on proving that the finite element scheme we apply is stable when c is replaced by its approximation. More precisely, we demonstrate that if c_h converge to c when h goes to zero (in a sense to be defined), the numerical solution converges to the analytical solution as both H and h go to zero.

We start by requiring approximation properties on the discretization space and we introduce the quantity $\mathcal{M}_{H,h}$ in definition 1 and 2. Note that the conditions given in definition 1 are fulfilled, for instance, by $\mathcal{P}_1$ Lagrangian polynomials.

Definition 1. *We consider a partition $\mathcal{T}_H$ of Ω. We assume that each cell $K \in \mathcal{T}_H$ is the image of a reference cell $\hat{K} \subset \mathbb{R}^2$ through an invertible affine map $\mathcal{F}_K \in \mathcal{L}(\mathbb{R}^2)$. We also consider a (finite dimensional) reference discretization space $\hat{P} \subset C^\infty(\hat{K})$, and define the discretization space V_H by*

$$V_H = \big\{ v_H \in H^1(\Omega, \mathbb{C}) \mid v_H|_K \circ \mathcal{F}_K \in \hat{P} \quad \forall K \in \mathcal{T}_H \big\}.$$

We further assume that there is a projection operator $\Pi_H \in \mathcal{L}(H^1(\Omega, \mathbb{C}), V_H)$ satisfying

$$|w - \Pi_H w|_{0,\Omega} \leq CH^2 |w|_{2,\Omega}, \quad |w - \Pi_H w|_{1,\Omega} \leq CH |w|_{2,\Omega}, \quad \forall w \in H^2(\Omega, \mathbb{C}),$$

where C is a constant depending on Ω, $\hat{K}$ and $\hat{P}$. Note that the multiplicative trace inequality ensures that

$$|w - \Pi_H w|_{0,\partial\Omega} \leq CH^{3/2} |w|_{2,\Omega},$$

where C is a constant depending on Ω, $\hat{K}$ and $\hat{P}$.

The construction of c_h is depicted at Section 4. In this section, assume that $c_h \in L^\infty(\Omega)$ and $c_{min} \leq c_h \leq c_{max}$. We also define the quantity $\mathcal{M}_{H,h}$:

Definition 2. *The velocity approximation error is defined by*

$$\mathcal{M}_{H,h} = \max_{K \in \mathcal{T}_H} \frac{1}{|K|} \int_K \left| \frac{1}{c^2} - \frac{1}{c_h^2} \right|,$$

where $|K|$ is the Lebesgue measure of the cell K.

In the following, we assume that $\mathcal{M}_{H,h}$ converges to zero as H and h go to zero. V_H and c_h being defined, we now introduce the discrete finite element problem. We write $k_h = \omega / c_h$. The discrete equation consists in finding $u_H \in V_H$ such that

$$B_h(u_H, v_H) = -\int_\Omega k_h^2 u_H \bar{v}_H - ik_{max} \int_{\partial\Omega} u_H \bar{v}_H + \int_\Omega \nabla u_H \cdot \nabla \bar{v}_H = \int_\Omega f \bar{v}_H, \quad \forall v_H \in V_H. \tag{9}$$

Proposition 3. *There exists a constant $C := C(c_{min}, \omega_0, \Omega)$ such that*

$$|B(u, v)| \leq C\big(\omega |u|_{0,\Omega} + |u|_{1,\Omega}\big)\big(\omega |v|_{0,\Omega} + |v|_{1,\Omega}\big), \quad \forall u, v \in H^1(\Omega, \mathbb{C}),$$

and

$$|B_h(u_H,v_H)| \leq C\big(\omega|u_H|_{0,\Omega}+|u_H|_{1,\Omega}\big)\big(\omega|v_H|_{0,\Omega}+|v_H|_{1,\Omega}\big), \quad \forall u_H, v_H \in V_H.$$

Proof. Since the proofs are similar for B and B_h, we focus on the first case only. Consider $u, v \in H^1(\Omega,\mathbb{C})$. It is obvious that

$$\begin{aligned} |B(u,v)| &\leq k_{max}^2|u|_{0,\Omega}|v|_{0,\Omega}+k_{max}|u|_{0,\partial\Omega}|v|_{0,\partial\Omega}+|u|_{1,\Omega}|v|_{1,\Omega} \\ &\leq \big(k_{max}|u|_{0,\Omega}+|u|_{1,\Omega}\big)\big(k_{max}|v|_{0,\Omega}+|v|_{1,\Omega}\big)+k_{max}|u|_{0,\partial\Omega}|v|_{0,\partial\Omega}. \end{aligned}$$

Moreover, for all $w \in H^1(\Omega,\mathcal{C})$, we have

$$\begin{aligned} k_{max}|w|_{0,\partial\Omega}^2 &\leq C(\Omega)k_{max}\big(|w|_{0,\Omega}^2+|w|_{0,\Omega}|w|_{1,\Omega}\big) \\ &\leq C(\Omega)k_{max}\big(|w|_{0,\Omega}^2+k_{max}|w|_{0,\Omega}^2+\frac{1}{k_{max}}|w|_{1,\Omega}^2\big) \\ &\leq C(\Omega,\omega_0,c_{min})\big(k_{max}^2|w|_{0,\Omega}^2+|w|_{1,\Omega}^2\big) \\ &\leq C(\Omega,\omega_0,c_{min})\big(k_{max}|w|_{0,\Omega}+|w|_{1,\Omega}\big)^2, \end{aligned}$$

and the result follows since $k_{max}=\omega/c_{min}$. □

We now give a result concerning the error induced by the approximation of the velocity parameter between the two sesquilinear forms B and B_h in Proposition 4.

Proposition 4. *There exists a constant $C := C(\hat{K},\hat{P})$ such that*

$$|B(u_H,v_H)-B_h(u_H,v_H)| \leq C\omega^2\mathcal{M}_{H,h}|u_H|_{0,\Omega}|v_H|_{0,\Omega}, \quad \forall u_H, v_H \in V_H.$$

Proof. Consider $u_H, v_H \in V_H$. We have

$$\begin{aligned} |B(u_H,v_H)-B_h(u_H,v_H)| &= \left|\int_\Omega (k^2-k_h^2)u_Hv_H\right| \\ &\leq \omega^2\sum_{K\in\mathcal{T}_H}\int_K\left|\frac{1}{c^2}-\frac{1}{c_h^2}\right||u_H||v_H| \\ &\leq \omega^2\sum_{K\in\mathcal{T}_H}|u_H|_{0,\infty,K}|v_H|_{0,\infty,K}\int_K\left|\frac{1}{c^2}-\frac{1}{c_h^2}\right|. \end{aligned} \tag{10}$$

Furthermore, for any cell $K \in \mathcal{T}_H$, $w_H \circ \mathcal{F}_K$ belongs to the finite dimensional space $\hat{P}$ if $w_H \in V_H$ and there exists a constant $\hat{C}$ depending on $\hat{P}$ only, such that

$$|w_H|_{0,\infty,K} = |w_H\circ\mathcal{F}_K|_{0,\infty,\hat{K}} \leq \hat{C}|w_H\circ\mathcal{F}_K|_{0,\hat{K}}.$$

We can thus derive

$$|w_H\circ\mathcal{F}_K|_{0,\hat{K}}^2 = \int_{\hat{K}}|w_H\circ\mathcal{F}_K|^2 = \text{Det } J_{\mathcal{F}_K}^{-1}\int_K|w_H|^2 = \frac{|\hat{K}|}{|K|}|w_H|_{0,K}^2,$$

so that

$$|w_H|_{0,\infty,K} \leq \hat{C}\sqrt{\frac{|\hat{K}|}{|K|}}|w_H|_{0,K}. \tag{11}$$

We can conclude by using (11) with $w_H = u_H, v_H$ in (10).

$$\begin{aligned} |B(u_H,v_H)-B_h(u_H,v_H)| &\leq \hat{C}^2|\hat{K}|\omega^2\sum_{K\in\mathcal{T}_H}\frac{|u|_{0,K}|v|_{0,K}}{|K|}\int_K\left|\frac{1}{c^2}-\frac{1}{c_h^2}\right| \\ &\leq \hat{C}^2|\hat{K}|\omega^2\mathcal{M}_{H,h}\sum_{K\in\mathcal{T}_H}|u|_{0,K}|v|_{0,K} \\ &\leq \hat{C}^2|\hat{K}|\omega^2\mathcal{M}_{H,h}|u|_{0,\Omega}|v|_{0,\Omega}. \end{aligned}$$

□

Before we establish our convergence result, we need three additional Lemma. In Lemma 2, we define the Ritz representation of the error z together with its elliptic projection z_H. We use the Ritz representation and its elliptic projection in Lemma 3 to bound the finite element error in the L^2 norm. Lemma 4 is a technical result required to prove the convergence in the H^1 norm in Theorem 2.

The proof of our error estimate is based on the theory of Zhu and Wu [25, 26] who establishes in particular Proposition 5.

In the remaining of this section $C := C(\Omega, c_{min}, c_{max}, x_0, \omega_0)$ denotes a constant independent of ω, H and h.

Proposition 5. *Let a be the sesquilinear form*

$$a(w,v) = \int_\Omega \nabla w \cdot \nabla \bar{v} - ik_{max} \int_{\partial\Omega} w\bar{v}, \quad \forall w, v \in H^1(\Omega, \mathbb{C}).$$

For all $z \in H^1(\Omega, \mathbb{C})$, there exist a unique $z_H \in V_H$ such that

$$a(w_H, z_H) = a(w_H, z), \quad \forall w_H \in V_H,$$

and we have

$$\begin{aligned} |z - z_H|_{0,\Omega} &\leq CH^2 |z|_{2,\Omega}, \\ |z - z_H|_{1,\Omega} &\leq CH |z|_{2,\Omega}, \\ |z - z_H|_{0,\partial\Omega} &\leq CH^{3/2} |z|_{2,\Omega}. \end{aligned}$$

Lemma 2. *Let $u_H \in V_H$ solve (9). Then there exists a unique element $z \in H^1(\Omega, \mathbb{C})$ such that,*

$$B(w,z) = \int_\Omega w\overline{u - u_H} \quad \forall w \in H^1(\Omega, \mathbb{C}),$$

and we have

$$|u - u_H|^2 = B(u - u_H, z). \tag{12}$$

Furthermore, there exists an element $z_H \in V_H$ such that

$$\frac{|B(u - u_H, z - z_H)|}{|u - u_H|_{0,\Omega}} \leq C\left(\omega^3 H^2 |u - u_H|_{0,\Omega} + \omega^2 H^2 |f|_{0,\Omega}\right). \tag{13}$$

Proof. According to Corollary 1, it is clear that there exist a unique $z \in H^1(\Omega, \mathbb{C})$ such that

$$B(w,z) = \int_\Omega w\overline{u - u_H}, \quad \forall w \in H^1(\Omega, \mathbb{C}).$$

In particular, picking $w = u - u_H$ yields (12).

Using Proposition 5, there exists an element $z_H \in V_H$ such that

$$a(u - u_H, z - z_H) = a(u - \Pi_H u, z - z_H)$$

It follows that

$$\begin{aligned} B(u - u_H, z - z_H) &= -\int_\Omega k^2 (u - u_H)\overline{(z - z_H)} + a(u - u_H, z - z_H) \\ &= -\int_\Omega k^2 (u - u_H)\overline{(z - z_H)} + a(u - \Pi_H u, z - z_H) \end{aligned}$$

Hence,

$$\begin{aligned} |B(u - u_H, z - z_H)| &\leq k_{max}^2 |u - u_H|_{0,\Omega} |z - z_H|_{0,\Omega} + k_{max} |u - \Pi_H u|_{0,\partial\Omega} |z - z_H|_{0,\partial\Omega} + |u - \Pi_H u|_{1,\Omega} |z - z_H|_{1,\Omega} \\ &\leq C\left(k_{max}^2 H^2 |u - u_H|_{0,\Omega} |z|_{2,\Omega} + k_{max} H^3 |u|_{2,\partial\Omega} |z|_{2,\partial\Omega} + H^2 |u|_{2,\Omega} |z|_{2,\Omega}\right) \\ &\leq C\left(\omega^2 H^2 |u - u_H|_{0,\Omega} |z|_{2,\Omega} + \omega H^3 |u|_{2,\Omega} |z|_{2,\Omega} + H^2 |u|_{2,\Omega} |z|_{2,\Omega}\right) \end{aligned}$$

Now, using Corollary 1 again, we have

$$|z|_{2,\Omega} \leq C\omega |u - u_H|_{0,\Omega},$$

and therefore

$$\frac{|B(u-u_H, z-z_H)|}{|u-u_H|_{0,\Omega}} \leq C\left(\omega^3 H^2 |u-u_H|_{0,\Omega} + \omega^2 H^3 |u|_{2,\Omega} + \omega H^2 |u|_{2,\Omega}\right).$$

We conclude thanks to Theorem 1. We have

$$|u|_{2,\Omega} \leq C\omega |f|_{0,\Omega},$$

and the proof follows since $\omega H \leq 1$. □

Lemma 3. *Let $u \in H^1(\Omega, \mathbb{C})$ solve (3) and let $u_H \in V_H$ be any solution to problem (9). Then if $\omega^3 H^2$ and $\omega \mathcal{M}_{H,h}$ are small enough, there exist a constant C such that*

$$|u-u_H|_{0,\Omega} \leq C\left(\omega^2 H^2 + \mathcal{M}_{H,h}\right)|f|_{0,\Omega}.$$

Proof. Recalling (12) from Lemma 2, there exits an element $z \in H^1(\Omega, \mathbb{C})$ such that

$$|u-u_H|^2_{0,\Omega} = B(u-u_H, z).$$

We then introduce $z_H \in V_H$ defined as in Lemma 2. Since u and u_H solve (3) and (9) respectively, we have

$$\begin{aligned} B(u-u_H, z) &= B(u-u_H, z-z_H) + B(u-u_H, z_H) \\ &= B(u-u_H, z-z_H) + B_h(u_H, z_H) - B(u_H, z_H), \end{aligned}$$

and therefore

$$|u-u_H|_{0,\Omega} \leq \frac{|B(u-u_H, z-z_H)|}{|u-u_H|_{0,\Omega}} + \frac{|B_h(u_H, z_H) - B(u_H, z_H)|}{|u-u_H|_{0,\Omega}}. \tag{14}$$

We bound the first term in the right hand side of (14) using Lemma 2. To deal with the second term, we recall Proposition 4: there holds

$$|B_h(u_H, z_H) - B(u_H, z_H)| \leq C\omega^2 \mathcal{M}_{H,h} |u_H|_{0,\Omega} |z_H|_{0,\Omega},$$

but we have

$$\begin{aligned} |z_H|_{0,\Omega} &\leq |z|_{0,\Omega} + |z-z_H|_{0,\Omega} \\ &\leq C\left(\omega^{-1}|u-u_H|_{0,\Omega} + H^2 |z|_{2,\Omega}\right) \\ &\leq C\left(\omega^{-1}|u-u_H|_{0,\Omega} + \omega H^2 |u-u_H|_{0,\Omega}\right) \\ &\leq C\omega^{-1}(1+\omega^2 H^2)|u-u_H|_{0,\Omega} \\ &\leq C\omega^{-1}|u-u_H|_{0,\Omega}, \end{aligned}$$

and

$$\begin{aligned} |u_H|_{0,\Omega} &\leq |u|_{0,\Omega} + |u-u_H|_{0,\Omega} \\ &\leq C\omega^{-1}|f|_{0,\Omega} + |u-u_H|_{0,\Omega}, \end{aligned}$$

so that

$$\frac{|B_h(u_H, z_H) - B(u_H, z_H)|}{|u-u_H|_{0,\Omega}} \leq C\left(\mathcal{M}_{H,h}|f|_{0,\Omega} + \omega \mathcal{M}_{H,h}|u-u_H|_{0,\Omega}\right).$$

Recalling (13) from Lemma 2, we obtain

$$|u-u_H|_{0,\Omega} \leq C\left(\omega^3 H^2 |u-u_H|_{0,\Omega} + \omega^2 H^2 |f|_{0,\Omega} + \mathcal{M}_{H,h}|f|_{0,\Omega} + \omega \mathcal{M}_{H,h}|u-u_H|_{0,\Omega}\right).$$

It follows that

$$\left(1 - C\omega^3 H^2 - C\omega \mathcal{M}_{H,h}\right)|u-u_H|_{0,\Omega} \leq C\left(\omega^2 H^2 + \mathcal{M}_{H,h}\right)|f|_{0,\Omega},$$

and we get Lemma 3 by assuming that $\omega^3 H^2$ and $\omega \mathcal{M}_{H,h}$ are small enough. □

Lemma 4. *The following estimate holds*

$$|u_H - \Pi_H u|^2_{1,\Omega} \leq C\big(\omega^2 |u-u_H|^2_{0,\Omega} + (\mathcal{M}^2_{H,h} + \omega^2 H^2)|f|^2_{0,\Omega}\big),$$

where u is the solution to (3) and u_H is any solution to (9).

Proof. First, the following relation holds

$$|u_H - \Pi_H u|^2_{1,\Omega} = \text{Re } B_h(u_H - \Pi_H u, u_H - \Pi_H u) + \int_\Omega k_h^2 |u_H - \Pi_H u|^2_{0,\Omega}. \tag{15}$$

Developing the first term of the right-hand-side in the above equation leads to:

$$\begin{aligned} B_h(u_H - \Pi_H u, u_H - \Pi_H u) &= B_h(u_H, u_H - \Pi_H u) - B_h(\Pi_H u, u_H - \Pi_H u) \\ &= B(u, u_H - \Pi_H u) - B_h(\Pi_H u, u_H - \Pi_H u) \\ &= B(u - \Pi_H u, u_H - \Pi_H u) + B(\Pi_H u, u_H - \Pi_H u) - B_h(\Pi_H u, u_H - \Pi_H u) \end{aligned}$$

It follows that

$$\begin{aligned} &\text{Re } B_h(u_H - \Pi_H u, u_H - \Pi_H u) \\ &\qquad \leq |B(u - \Pi_H u, u_H - \Pi_H u)| + |B(\Pi_H u, u_H - \Pi_H u) - B_h(\Pi_H u, u_H - \Pi_H u)|. \end{aligned} \tag{16}$$

Then, using Proposition 3 and Theorem 1, we have:

$$\begin{aligned} |B(u - \Pi_H u, u_H - \Pi_H u)| &\leq C\big(\omega |u - \Pi_H u|_{0,\Omega} + |u - \Pi_H u|_{1,\Omega}\big)^2 \\ &\leq C\big(\omega^2 |u - \Pi_H u|^2_{0,\Omega} + |u - \Pi_H u|^2_{1,\Omega}\big) \\ &\leq C\big(\omega^2 H^2 |u|_{2,\Omega} |u_H - \Pi_H u|_{0,\Omega} + H |u|_{2,\Omega} |u_H - \Pi_H u|_{1,\Omega}\big) \\ &\leq C\big(\omega^3 H^2 |f|_{0,\Omega} |u_H - \Pi_H u|_{0,\Omega} + \omega H |f|_{0,\Omega} |u_H - \Pi_H u|_{1,\Omega}\big) \\ &\leq C\big(\omega^2 H^2 |f|^2_{0,\Omega} + \omega^4 H^2 |u_H - \Pi_H u|^2_{0,\Omega} + \frac{\omega^2 H^2}{\eta} |f|^2_{0,\Omega} + \eta |u_H - \Pi_H u|^2_{1,\Omega}\big) \\ &\leq \frac{1}{2} |u_H - \Pi_H u|^2_{1,\Omega} + C\big(\omega^4 H^2 |u_H - \Pi_H u|^2_{0,\Omega} + \omega^2 H^2 |f|^2_{0,\Omega}\big). \end{aligned} \tag{17}$$

Moreover, Proposition 4 implies that

$$\begin{aligned} |B(\Pi_H u, u_H - \Pi_H u) - B_h(\Pi_H u, u_H - \Pi_H u)| &\leq C\omega^2 \mathcal{M}_{H,h} |\Pi_H u|_{0,\Omega} |u_H - \Pi_H u|_{0,\Omega} \\ &\leq C\omega^2 \big(\mathcal{M}^2_{H,h} |\Pi_H u|^2_{0,\Omega} + |u_H - \Pi_H u|^2_{0,\Omega}\big) \\ &\leq C\omega^2 \big(\mathcal{M}^2_{H,h} (|u|^2_{0,\Omega} + |u - \Pi_H u|^2_{0,\Omega}) + |u_H - \Pi_H u|^2_{0,\Omega}\big) \\ &\leq C\omega^2 \big(\mathcal{M}^2_{H,h} (|u|^2_{0,\Omega} + H^4 |u|^2_{2,\Omega}) + |u_H - \Pi_H u|^2_{0,\Omega}\big) \\ &\leq C\omega^2 \big(\mathcal{M}^2_{H,h} (\frac{1}{\omega^2} + \omega^2 H^4) |f|^2_{0,\Omega} + |u_H - \Pi_H u|^2_{0,\Omega}\big) \\ &\leq C\big(\mathcal{M}^2_{H,h} (1 + \omega^4 H^4) |f|^2_{0,\Omega} + \omega^2 |u_H - \Pi_H u|^2_{0,\Omega}\big). \end{aligned} \tag{18}$$

Now, since that $k_h^2 \leq C\omega^2$, we have

$$|u_H - \Pi_H u|^2_{1,\Omega} \leq \text{Re } B_h(u_H - \Pi_H u, u_H - \Pi_H u) + C\omega^2 |u_H - \Pi_H u|^2_{0,\Omega}$$

Plugging (17) and (18) in (16) implies that

$$\frac{1}{2} |u_H - \Pi_H u|^2_{1,\Omega} \leq C\big\{(\omega^2 + \omega^4 H^2) |u_H - \Pi_H u|^2_{0,\Omega} + (\mathcal{M}^2_{H,h} (1 + \omega^4 H^4) + \omega^2 H^2) |f|^2_{0,\Omega}\big\}.$$

Then if ωH is small enough, we end up with

$$|u_H - \Pi_H u|^2_{1,\Omega} \leq C\big\{\omega^2 |u_H - \Pi_H u|^2_{0,\Omega} + (\mathcal{M}^2_{H,h} + \omega^2 H^2) |f|^2_{0,\Omega}\big\}.$$

We end the demonstration by observing that

$$\begin{aligned} \omega^2 |u_H - \Pi_H u|^2_{0,\Omega} &\leq \omega^2 |u - u_H|^2_{0,\Omega} + \omega^2 |u - \Pi_H u|^2_{0,\Omega} \\ &\leq \omega^2 |u - u_H|^2_{0,\Omega} + C\omega^2 H^4 |u|^2_{2,\Omega} \\ &\leq \omega^2 |u - u_H|^2_{0,\Omega} + C\omega^4 H^4 |f|^2_{0,\Omega}. \end{aligned}$$

□

We now establish a convergence result under the assumption that $\omega^3 H^2$ and $\omega \mathcal{M}_{H,h}$ can be made arbitrarily small.

Theorem 2. *Assume that $\omega^3 H^2$ and $\omega\mathcal{M}_{H,h}$ are small enough. Then problem (9) has a unique solution $u_H \in V_H$. Furthermore, u_H satisfies*

$$\omega|u-u_H|_{0,\Omega} + |u-u_H|_{1,\Omega} \leq C\left(\omega\mathcal{M}_{H,h} + \omega H + \omega^3 H^2\right)|f|_{0,\Omega}, \tag{19}$$

where $C := C(\Omega, c_{min}, c_{max}, x_0, \omega_0)$ denotes a constant independent of ω, H and h.

Proof. Let us first show existence end uniqueness of u_H. Since V_H is a finite dimensional space, (9) is equivalent to a linear system with size $\left(\dim V_H \times \dim V_H\right)$. Therefore, we only need to prove uniqueness. Assume then that $f = 0$ in the discrete and continuous problem (3) and (9). According to Theorem 1, the corresponding continuous solution u is $u = 0$. Then, from Theorem 2, we deduce that

$$|u_H|_{0,\Omega} \leq C\omega^2 H^2 |f|_{0,\Omega} = 0,$$

so that $u_H = 0$ and uniqueness occurs.

We now turn to the proof of error estimate (19). Recalling lemma 3, it is clear that

$$\omega|u-u_H|_{0,\Omega} \leq C\left(\omega^3 H^2 + \omega\mathcal{M}_{H,h}\right)|f|_{0,\Omega},$$

and it remains to show that

$$|u-u_H|_{1,\Omega} \leq C\left(\omega\mathcal{M}_{H,h} + \omega H + \omega^3 H^2\right)|f|_{0,\Omega}.$$

To start with, it is clear that

$$\begin{aligned} |u-u_H|_{1,\Omega} &\leq |u-\Pi_H u|_{1,\Omega} + |u_H - \Pi_H u|_{1,\Omega} \\ &\leq CH|u|_{2,\Omega} + |u_H - \Pi_H u|_{1,\Omega} \\ &\leq C\omega H|f|_{0,\Omega} + |u_H - \Pi_H u|_{1,\Omega}, \end{aligned}$$

but recalling Lemma 4, we have

$$\begin{aligned} |u_H - \Pi_H u|^2 &\leq C\left(\omega^2|u-u_H|^2_{0,\Omega} + (\mathcal{M}^2_{H,h} + \omega^2 H^2)|f|_{0,\Omega}\right) \\ &\leq C\left(\omega^6 H^4 + \mathcal{M}^2_{H,h} + \omega^2 H^2\right)|f|^2_{0,\Omega}, \end{aligned}$$

hence

$$|u_H - \Pi_H u| \leq C\left(\omega^3 H^2 + \mathcal{M}_{H,h} + \omega H\right)|f|_{0,\Omega},$$

and the result follows since $\mathcal{M}_{H,h} \leq \omega_0^{-1}\omega\mathcal{M}_{H,h}$. □

3. Approximation of c

In this section we discuss how to pick an approximation c_h of c which is both accurate and easy to compute. Regarding the accuracy, we propose to quantify it by the measurement of $\mathcal{M}_{H,h}$ previously introduced at Definition 2. For the sake of simplicity, we restrict our study to the case of flat interfaces. We also show that the entries of the linear system related to (9) are easy to compute.

The approximation of c is based upon the following procedure. Let $\mathcal{T}_h$ be a given partition of the reference cell $\hat{K}$. We can map this partition to each actual cell $K \in \mathcal{T}_H$ and thus obtain a partition $\mathcal{T}^K_{H,h}$ of the cell K. Finally, gathering all the partitions associated to each cell $K \in \mathcal{T}_H$ together, we obtain a (possibly non-conforming) partition $\mathcal{T}_{H,h}$ of Ω (see Figure 3 which illustrates this process). The approximate velocity parameter is defined as follow:

Definition 3. *Let $c \in L^\infty(\Omega)$ be the global velocity supposed to satisfy assumption 2. Let $x_A \in A$ be the barycenter of $A \in \mathcal{T}_{H,h}$. If x_A does not belong to an interface, we set $c_h|_A = c(x_A)$, otherwise we define $c_h|_A = \sup_A c$.*

Our definition of c_h corresponds to a P_0-interpolation of c. Recalling Definition 2, it is clear that other choices are possible and covered by our convergence analysis. However we consider P_0-interpolation only. Indeed, since we consider a piecewise constant parameters, it is not clear that higher order approximations might bring additional precision. Furthermore, difficulties can arise when defining high order approximation of c. For instance, it is shown in [20] that if c_h can take negative values if it is defined as a P_2-interpolation of c.

Remark 1. *As indicated in the end of this section, a high number of subcells can be considered in the partition $\mathcal{T}_h$ without changing the structure of the finite-element linear system. Thus, it is possible to use a sufficiently fine h so that the medium is "well represented". We further point out that if needed, an upper bound of the constant $\mathcal{M}_{H,h}$ could be estimated numerically for a given finite-element mesh $\mathcal{T}_H$, leading to a practical estimation of the required h.*

We now show that in the simple case of flat interfaces, the quantity $\mathcal{M}_{H,h}$ goes to zero as h goes to zero uniformly with respect to H. Figure 3 is helpful to figure out different quantities used in the demonstration.

Proposition 6. *Assume that the interfaces of the partition (Ω_r) are flat and that the medium approximation submesh $\mathcal{T}_h$ is regular. Then there exists a constant C depending on the reference cell $|\hat{K}|$ only such that*

$$\mathcal{M}_{H,h} \leq CRh \left| \frac{1}{c_{min}^2} - \frac{1}{c_{max}^2} \right|.$$

Proof. Consider a given cell $K \in \mathcal{T}_H$. Then, K is crossed by at-most R straight interfaces and, since the submesh $\mathcal{T}_{H,h}^K$ is regular, there exists a constant C such that the number of subcells $A \in \mathcal{T}_{H,h}^K$ crossed by each interface is less than C/h. Then the total number of subcells of K crossed by an interface is less than CR/h.

We can easily upper-bound the measure $|A|$ of each subcell $A \in \mathcal{T}_h$ like $|A| \leq Ch^2$. Since the submesh $\mathcal{T}_{H,h}^K$ is constructed from a linear mapping, it follows that for all $A \in \mathcal{T}_{H,h}^K$

$$|A| \leq C \frac{|K|}{|\hat{K}|} h^2.$$

Let $\mathcal{A}_c \subset \mathcal{T}_{H,h}^K$ be the set of all subcells crossed by an interface. The total measure of the crossed subcells is then satisfying

$$\sum_{A \in \mathcal{A}_c} |A| \leq CR \frac{|K|}{|\hat{K}|} h.$$

Next, let $\mathcal{A}_e = \mathcal{T}_{H,h}^K \setminus \mathcal{A}_c$ be the set of subcells which are not crossed by any interface. Then, the approximation of c by c_h is exact on each cell $A \in \mathcal{A}_e$. Therefore, we have

$$\int_K \left| \frac{1}{c^2} - \frac{1}{c_h^2} \right| \leq \sum_{A \in \mathcal{A}_c} \int_A \left| \frac{1}{c^2} - \frac{1}{c_h^2} \right| \leq \left| \frac{1}{c_{min}^2} - \frac{1}{c_{max}^2} \right| \sum_{A \in \mathcal{A}_c} |A| \leq CR \frac{|K|}{|\hat{K}|} h \left| \frac{1}{c_{min}^2} - \frac{1}{c_{max}^2} \right|.$$

But by definition of $\mathcal{M}_{H,h}$, we have

$$\mathcal{M}_{H,h} = \max_{K \in \mathcal{T}_{H,h}} \frac{1}{|K|} \int_K \left| \frac{1}{c^2} - \frac{1}{c_h^2} \right| \leq \frac{C}{|\hat{K}|} Rh \left| \frac{1}{c_{min}^2} - \frac{1}{c_{max}^2} \right|.$$

which concludes the proof of Proposition 6. □

To end up with this section, we discuss on the computational cost of the proposed method. The corresponding linear system reads nearly as the one related to the classical FEM, except that the coefficients of the discrete system are weighted differently just because c_h is different. Therefore, only the construction of the linear system is more expensive. To compute the entries of the linear system, we first compute reference integrals on each subcell $B \in \mathcal{T}_h$. This is done once and for all at the beginning of the simulation (or directly hard-coded, if the mesh $\mathcal{T}_h$ is known before execution) and it corresponds thus to a pre-processing step. Next, the mapping $\mathcal{F}_K$ is used to compute the coefficients associated with each cell K.

Let $\{\hat{\varphi}_i\}_{i=1}^D$ be a basis of $\hat{P}$. Note that if $\hat{P} = \mathcal{P}_k(\hat{K})$, $D = p(p+1)/2$. On each cell $K \in \mathcal{T}_H$, one has to compute

$$\int_K k_h^2 \hat{\varphi}_i \circ \mathcal{F}_K^{-1} \varphi_j \circ \mathcal{F}_K^{-1} = \sum_{A \in \mathcal{T}_{H,h}^K} k_h^2 \int_A \hat{\varphi}_i \circ \mathcal{F}_K^{-1} \varphi_j \circ \mathcal{F}_K^{-1} = \text{Det } J_{\mathcal{F}_K} \sum_{B \in \mathcal{T}_h} k_h^2 \int_B \hat{\varphi}_i \hat{\varphi}_j.$$

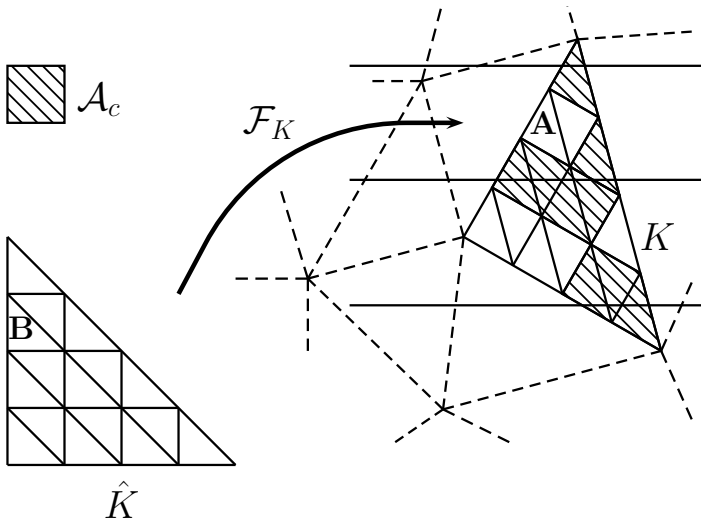


FIGURE 3. Mapping of the reference submesh

It should be noted that the last integral is independent of the given cell K. Therefore, we may compute the reference integrals

$$M_{ij}^B = \int_B \hat{\varphi}_i \hat{\varphi}_j, \quad \forall B \in \mathcal{T}_h.$$

once and for all independently of the number of coarse cells. The corresponding computational cost is thus insignificant. Now, for a given cell K, we have to compute

$$\text{Det } J_{\mathcal{F}_K} \sum_{B \in \mathcal{T}_h} k_h^2 M_{ij}^B.$$

If N_h is the number of cell in $\mathcal{T}_h$, we thus need to perform N_h multiplications, $N_h - 1$ additions, and one multiplication by the Jacobian, which comes to $2N_h$ operations for each coefficient. Now, arguing the symmetry of the system, we only need to compute $D(D+1)/2$ coefficients, which requires $N_h D(D+1)$ operations per cell. Then, if we assume that the mesh $\mathcal{T}_h$ is regular, $N_h \leq C/h^2$ and the number of operations per cell is of $\mathcal{O}(D(D+1)/h^2)$ operations.Another way to think about it, is that if we are using N_h subcells, the computational cost of the matrix assembly is multiplied by N_h. Note that only the cost of the assembly is increased, since the linear system keeps the same size and stencil.

4. Numerical experiments

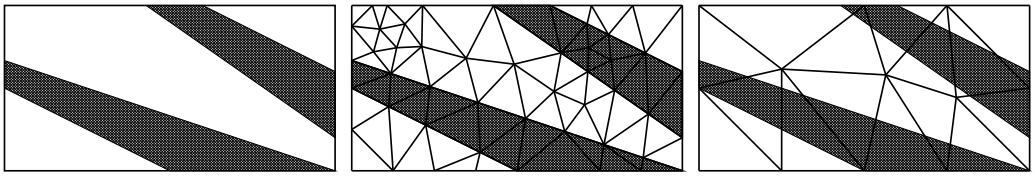
The objective of this section is to deliver performance assessments of the MMAm. We base our analysis on artificial stratified media in which we have an analytical solution. In particular, we illustrate how the MMAm performs well even when the velocity is strongly varying and does not satisfy the technical assumption 2. The performance of the method is measured from the values of the $L^2(\Omega)$ norm relative error, that is

$$E = \frac{\int_\Omega |u - u_{H,h}|^2 dx}{\int_\Omega |u|^2 dx} \tag{20}$$

where u denotes the exact (analytical) solution and $u_{H,h}$ is the numerical solution.

The numerical results are depicted by the mean of the solution profile, that is the graph of $x_2 \to u_{H,h}(500, x_2)$.

All along this section, we use two kinds of meshes as depicted in figure 4. Some are constructed so that the velocity is constant inside each cell. We then speak about fitting meshes in contrast to non-fitting meshes which are composed of cells inside which the velocity may vary. Obviously,



(a) (b) (c)

FIGURE 4. Velocity model (a), fitting (b) and non-fitting (c) meshes

the MMAm must be used on non-fitting meshes to take into account subcells velocity variations. Standard FEM, or other usual methods, are rather used on fitting meshes.

Herein, we will also consider the standard FEM on non-fitting meshes. In this case, we transform the velocity parameter so that it is constant in each cell of the mesh. We use two different strategies. The first idea is to select the value of the velocity parameter in the center of the cell. It corresponds to using the MMAm with only one subcell. The other strategy is to average the velocity parameter on the cell and choosing the value

$$\frac{1}{c_K^2} = \int_K \frac{1}{c^2}.$$

When analysing MMAm results, we will distinguish between the FEM approximation error and the medium approximation error. The FEM approximation error is defined as the error of the best approximation, i.e.

$$E_{FEM} = \inf_{v_H \in V_H} |u - v_H|_{0,\Omega},$$

while the medium approximation error is defined as $E_{MED} = \mathcal{M}_{H,h}$. We observe that for a given mesh (i.e. H is fixed), the FEM approximation error is fixed but the medium approximation error can be reduced by refining the submesh (i.e. h goes to zero).

In each of the following examples, we consider a fixed propagation medium together with a given mesh and an approximation order. We present the results obtained for different values of ω and h. In particular, we show that in the case where the dominant part of the error is due to the medium approximation, the quality of the numerical solution can be slightly improved by increasing the number of subcells.

For the computations, we use triangular Lagrangian finite elements. The medium approximation submesh is obtained through a homothety of the reference triangle, as shown in Figure 5. Note that those meshes are obviously regular and satisfy the hypothesis of Proposition 6.

4.1. **Analytical solution.** To construct an analytical solution, we introduce an auxiliary 1D problem, that is to find $u \in C^1([0, L])$ such that

$$\begin{cases} -\dfrac{\omega^2}{c^2(x)} v(x) - v''(x) &= 0 \quad \text{for } x \in [0, L] \\ -v'(0) &= 1 \\ v'(1) - i\dfrac{\omega}{c(1)} v(1) &= 0, \end{cases}$$

where c is piecewise constant on a partition $0 = x_0 < x_1 < \cdots < x_m = L$. Its unique solution is then given by

$$v|_{[x_{j-1},x_j]}(x) = \alpha_j^i e^{i\omega/c_i x} + \alpha_j^r e^{-i\omega/c_i x}, \quad v|_{[x_{L-1},x_L]}(x) = \alpha_L^i e^{i\omega/c_L x},$$

where the coefficients α are computed by solving a linear system given by the C^1 compatibility conditions at each point x_j with $0 < j < m$ and the condition $-v'(0) = 1$. We then get a two dimensional problem by setting $\Omega = (0, 1000) \times (0, L)$ and $k \in L^\infty(\Omega)$ is defined as $k \in L^\infty(\Omega)$,

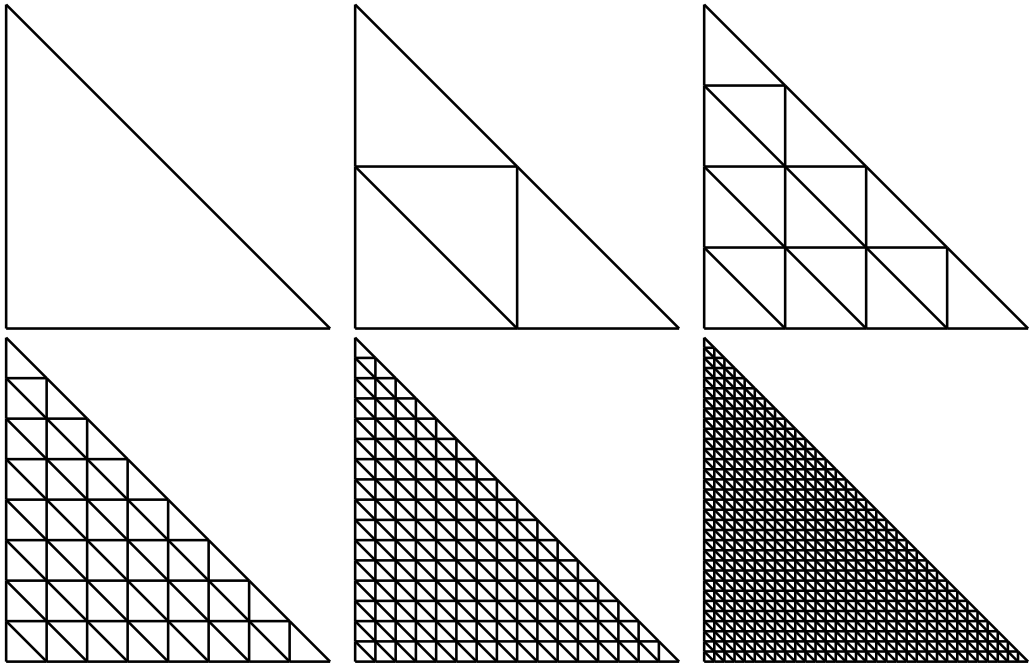


FIGURE 5. Velocity approximation schemes for $h = 1, 0.5, 0.25, 0.125, 0.0625$ and 0.03125.

$k|_{\Omega_j} = \omega/c_j$ where $\Omega_j = (0, 1000) \times (x_{j-1}, x_j)$. Then $u(x_1, x_2) = v(x_2)$ is the unique solution to

$$\begin{cases} -k^2 u - \Delta u &= 0 & \text{in } \Omega \\ \partial_{\mathbf{n}} u &= 1 & \text{on } (0,1000) \times \{0\} \\ \partial_{\mathbf{n}} u - ik_L u &= 0 & \text{on } (0,1000) \times \{L\} \\ \partial_{\mathbf{n}} u &= 0 & \text{on } \{0\} \times (0, L) \\ \partial_{\mathbf{n}} u &= 0 & \text{on } \{1000\} \times (0, L). \end{cases}$$

4.2. **A two-layered media.** We begin with evaluating the medium approximation error as a function of h. For that purpose, we consider the case of a two-layered medium composed of two homogeneous layers. In this case, the use of a fitting mesh is obviously relevant and this case gives us a way to measure the effect of MMAm on the accuracy of the solution.

We set $x_0 = 0, x_1 = 500, x_2 = L = 1000$, $c_1 = 1000$ and $c_2 = 2000$. In order to quantify the error coming from the medium approximation we use both a fitting and a non-fitting meshes. When using the fitting mesh, the medium is perfectly represented, since the coefficient c is constant in each cell of the finite element mesh. On the other hand, when using the non-fitting mesh, c must be approximated by c_h since it may vary inside an element. The experiment then shows that when the velocity approximation is refined, the solution error obtained with the non-fitting mesh is getting closer to the error obtained on the fitting mesh.

The non-fitting mesh contains 164 cells and the fitting mesh contains 166 cells. We start with $\mathcal{P}_2$ elements and the corresponding results are represented in the Table 1.

In the first column, the integer numbers indicate the number of subcells that are used to approximate the velocity inside each cell of the non-fitting mesh. The last line stands for the results obtained by using the standard $\mathcal{P}_2$ FEM with the fitting mesh.

We can observe that for each value of ω, the error decreases when letting h go to 0. Moreover, when comparing with the last line of the table, we can see that the MMAm reaches the same level of accuracy that the standard $\mathcal{P}_2$ FEM. When the frequency is increasing, the two methods result in the same level of accuracy and MMAm accuracy seems to reach a plateau. We believe that the medium approximation error becomes so small that quickly the values of the error describe the finite element approximation only.

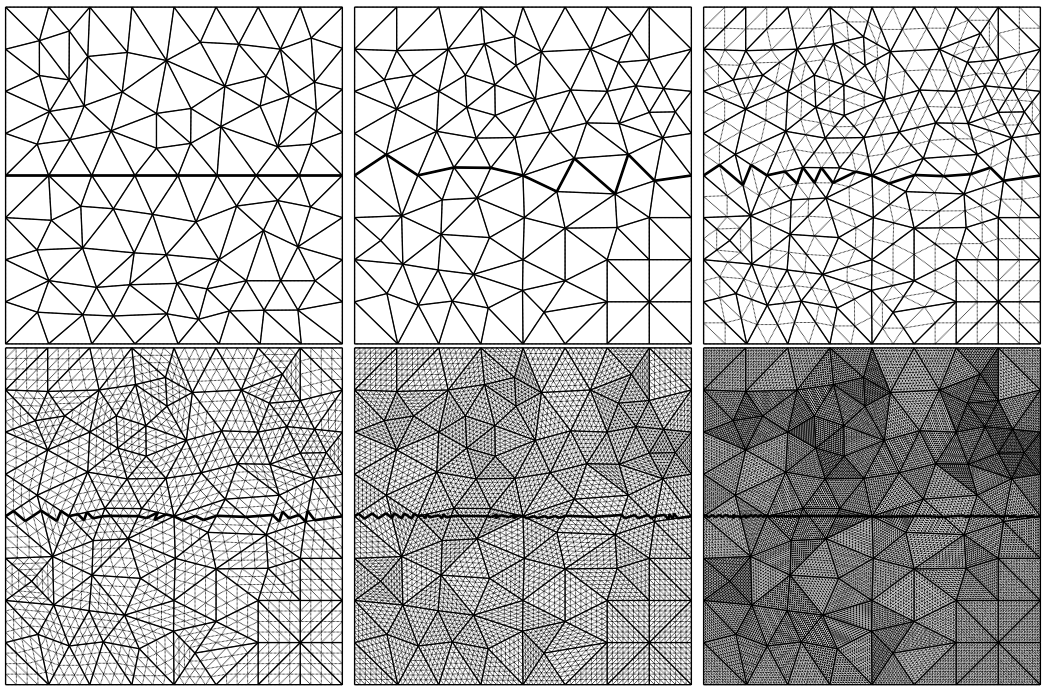


FIGURE 6. Evolution of the interface: fitting mesh (top-left) and non fitting mesh with $h = 1, 0.5, 0.25, 0.125$ and 0.0625.

$\mathcal{P}_2$	$\omega = 2\pi$	$\omega = 4\pi$	$\omega = 6\pi$
1	9.76×10^{-2}	2.38×10^{-1}	9.11×10^{-1}
4	2.26×10^{-2}	7.92×10^{-2}	3.24×10^{-1}
16	1.18×10^{-2}	4.62×10^{-2}	2.02×10^{-1}
64	5.20×10^{-3}	3.76×10^{-2}	2.05×10^{-1}
256	3.05×10^{-3}	3.61×10^{-2}	2.09×10^{-1}
1024	2.59×10^{-3}	3.59×10^{-2}	2.11×10^{-1}
fitting	1.81×10^{-3}	3.78×10^{-2}	2.65×10^{-1}

TABLE 1. $\mathcal{P}_2$ elements

$\mathcal{P}_4$	$\omega = 2\pi$	$\omega = 4\pi$	$\omega = 6\pi$	$\omega = 8\pi$	$\omega = 10\pi$
1	9.67×10^{-2}	2.25×10^{-1}	3.42×10^{-1}	4.81×10^{-1}	5.01×10^{-1}
4	2.22×10^{-2}	6.59×10^{-2}	1.42×10^{-1}	4.03×10^{-1}	1.90×10^{-1}
16	1.22×10^{-2}	3.75×10^{-2}	6.65×10^{-2}	2.37×10^{-1}	8.94×10^{-2}
64	4.70×10^{-3}	1.44×10^{-2}	2.74×10^{-2}	9.81×10^{-2}	4.50×10^{-2}
256	1.47×10^{-3}	4.91×10^{-3}	1.13×10^{-2}	4.54×10^{-2}	2.94×10^{-2}
1024	5.25×10^{-4}	1.54×10^{-3}	4.58×10^{-3}	1.67×10^{-2}	2.52×10^{-2}
fitting	2.62×10^{-6}	8.80×10^{-5}	8.10×10^{-4}	5.76×10^{-3}	2.44×10^{-2}

TABLE 2. $\mathcal{P}_4$ elements

Table 2 represents the results obtained when using $\mathcal{P}_4$ elements. The same conclusions hold except that due to a highest degree of approximation, the medium approximation error stabilizes itself on a plateau for $\omega = 10\pi$ only. It is worth noting than when ω is less than 10π, the convergence is super linear which illustrates well Section 3 results.

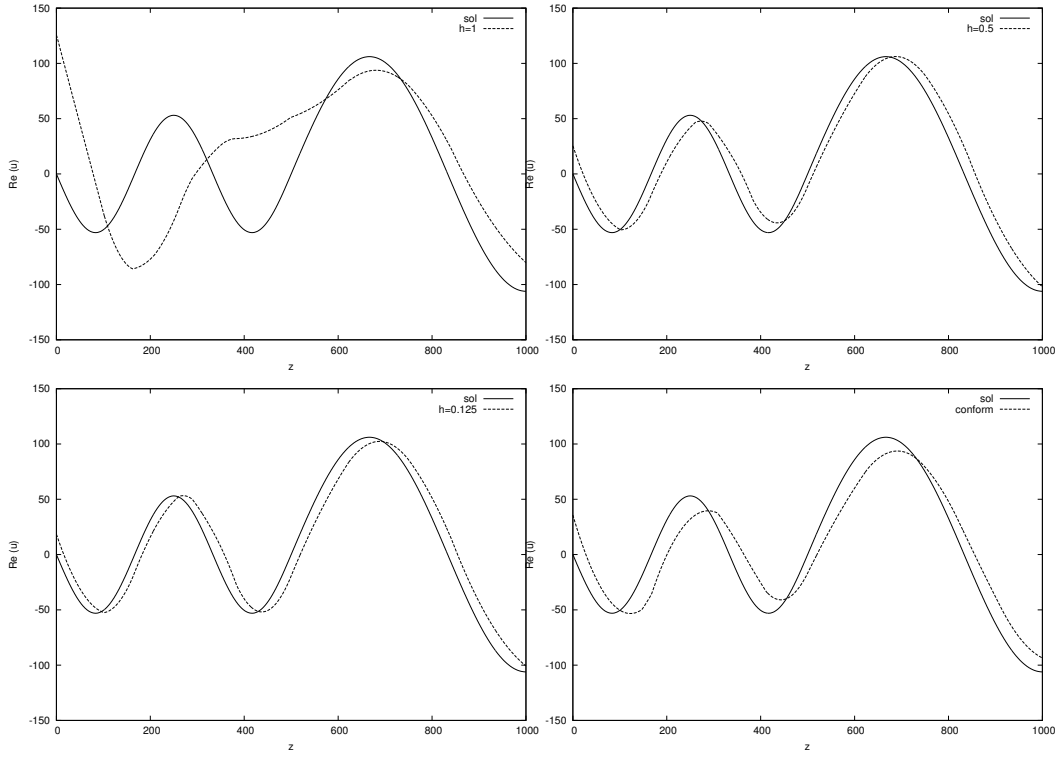


FIGURE 7. Solution profile for $\mathcal{P}_2$ elements, $\omega = 6\pi$

$\mathcal{P}_6$	$\omega = 20\pi$	$\omega = 30\pi$	$\omega = 40\pi$	$\omega = 50\pi$	$\omega = 60\pi$
mean	4.38×10^{-2}	1.70×10^{-1}	6.44×10^{-1}	1.90×10^{-1}	2.33×10^{0}
1	4.19×10^{-2}	1.61×10^{-1}	5.04×10^{-1}	1.87×10^{-1}	1.19×10^{0}
4	7.27×10^{-3}	2.39×10^{-2}	4.83×10^{-1}	1.02×10^{-1}	4.47×10^{-1}
16	2.12×10^{-3}	7.06×10^{-3}	5.97×10^{-2}	6.63×10^{-2}	3.52×10^{-1}
64	1.02×10^{-3}	3.76×10^{-3}	3.64×10^{-2}	6.33×10^{-2}	3.34×10^{-1}
256	4.93×10^{-4}	1.74×10^{-3}	3.52×10^{-2}	6.26×10^{-2}	3.40×10^{-1}
1024	2.00×10^{-4}	9.40×10^{-4}	3.69×10^{-2}	6.19×10^{-2}	3.37×10^{-1}

TABLE 3. Multi-layered medium

4.3. **Multi-layered medium.** We now set $L = 3000$. We decompose the propagation domain into 1000 layers of 3 meters each. We set $c_{min} = 1500$, $c_{max} = 5500$. The velocity parameter varies linearly from $c_1 = c_{min}$ to $c_{1000} = c_{max}$. We use $\mathcal{P}_6$ elements on a 1033 cells mesh. We carry out simulations for different values of h. To compare with parameter averaging methods, we perform simulations for $k_h^2|_K$ given as the mean value of k^2 on the cell K.

On table 3, we present the results that we have obtained by discretizing with $\mathcal{P}_6$ Lagrangian elements. We can draw the same conclusion than in the previous test case. It is interesting to note that the MMAm results are always better then when the standard FEM is used with the mean value of the wavenumber in each cell. This example shows that the subscheme quadrature strategy of the MMAm is superior to a simple averaging of the wavenumber, as depicted by the first line of Table 3.

It is also clear that for a given pulsation, reducing the approximation step h reduces the solution error. For the lowest pulsation $\omega = 20\pi$, the convergence is super linear, which is consistent with the results of section 3. For higher pulsations, the part of the error due to finite element approximation is much larger, so that the linear convergence is not observed anymore. For low

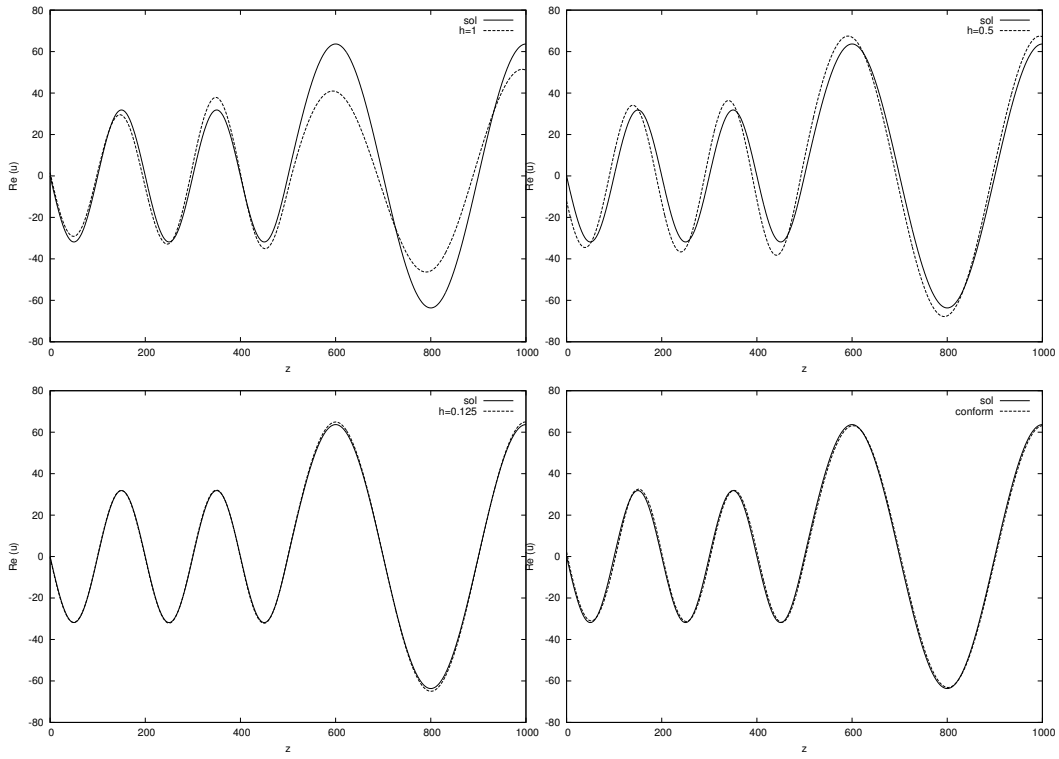


FIGURE 8. Solution profile for $\mathcal{P}_4$ elements, $\omega = 10\pi$

$\mathcal{P}_6$	$\omega = 20\pi$	$\omega = 30\pi$	$\omega = 40\pi$	$\omega = 50\pi$
mean	1.06×10^{0}	6.73×10^{-1}	1.17×10^{0}	2.76×10^{0}
1	9.99×10^{-1}	1.81×10^{0}	7.44×10^{0}	3.20×10^{0}
4	7.41×10^{-1}	3.84×10^{0}	1.71×10^{0}	1.88×10^{0}
16	3.41×10^{-1}	6.79×10^{-1}	3.34×10^{0}	2.68×10^{0}
64	3.12×10^{0}	1.86×10^{-1}	4.45×10^{-1}	1.05×10^{0}
256	8.40×10^{-2}	6.60×10^{-2}	1.03×10^{-1}	2.77×10^{-1}
1024	6.23×10^{-2}	3.63×10^{-2}	7.00×10^{-2}	2.12×10^{-1}

TABLE 4. Highly heterogeneous multi-layered medium

frequency regime, the solution is accurate for $h = 1$. This illustrates 6 which states that higher pulsations are more sensitive to the accuracy of the medium approximation.

4.4. **Multi-layered medium: Highly heterogeneous.** We consider here the case where the velocity does not satisfy the technical condition (2). The velocity model is now constructed by modifying the previous one as follows. Between 0 and 1500 meters and between 2000 and 3000 meters, the velocity is decreased by 500 one layer from two, and increased by 500 in the remaining layers. Between 1500 and 2000 meters, the velocity is 500 in one layer from two. We use an adaptive mesh, which is more refined between 1500 and 2000 meters in order to correctly fit the small wavelength in this area. The mesh is made of 4838 cells and is represented in Figure 10.

On figure 11, we have plotted the solution profile and we observe that the MMAm solution is accurate as soon as h is less that 0.0625, which means that we need to use at least 256 subcells to compute the entries of the matrix. This is not surprising because we consider a velocity model including very strong contrasts. It is indeed composed of very thin layers and the variations of the velocity are so important that the averaging technique completely fails.

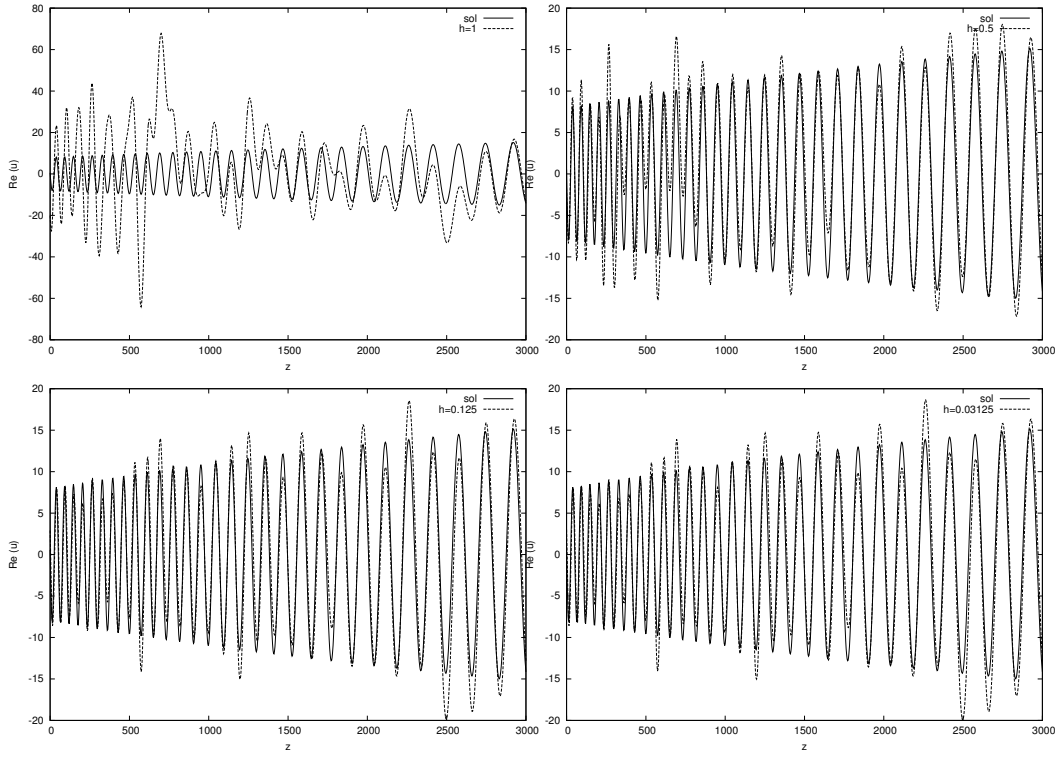


FIGURE 9. Solution profile in gradient domain for $\mathcal{P}_6$ elements, $\omega = 60\pi$

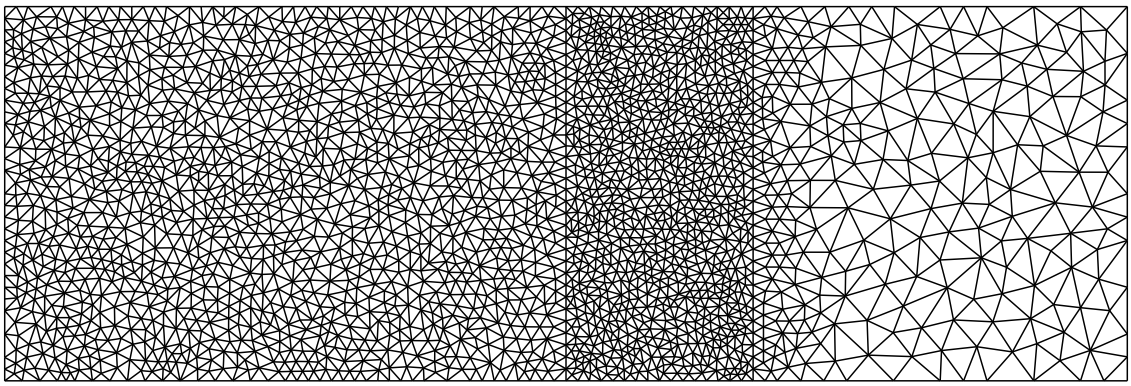


FIGURE 10. Adaptive mesh (90 degrees rotation)

4.5. **High order MMAm VS fitting mesh based method.** In the previous numerical tests, we have validated the MMAm and we have concluded that when using enough subcells we obtain accurate results even in highly heterogeneous media. In particular, the first experiment showed that when a fitting mesh is available, the accuracy of the MMAm on a non-fitting mesh of the same size is comparable to the standard FEM on the fitting mesh.

In this section, we investigate the reduction of the computational cost offered by the MMAm to obtain a 5% relative error on the previous velocity model at the frequency $\omega = 40\pi$.

We use regular meshes based on cartesian grids of different sizes and different polynomial degrees. As a starting point, we discretize the problem with the coarsest possible fitting mesh. The mesh steps are given by $h_x = 3.33m$, $h_z = 3m$. The z step is chosen to be exactly the length of a layer, so that the mesh is fitting, and the x step is chosen so that the grid cells are nearly

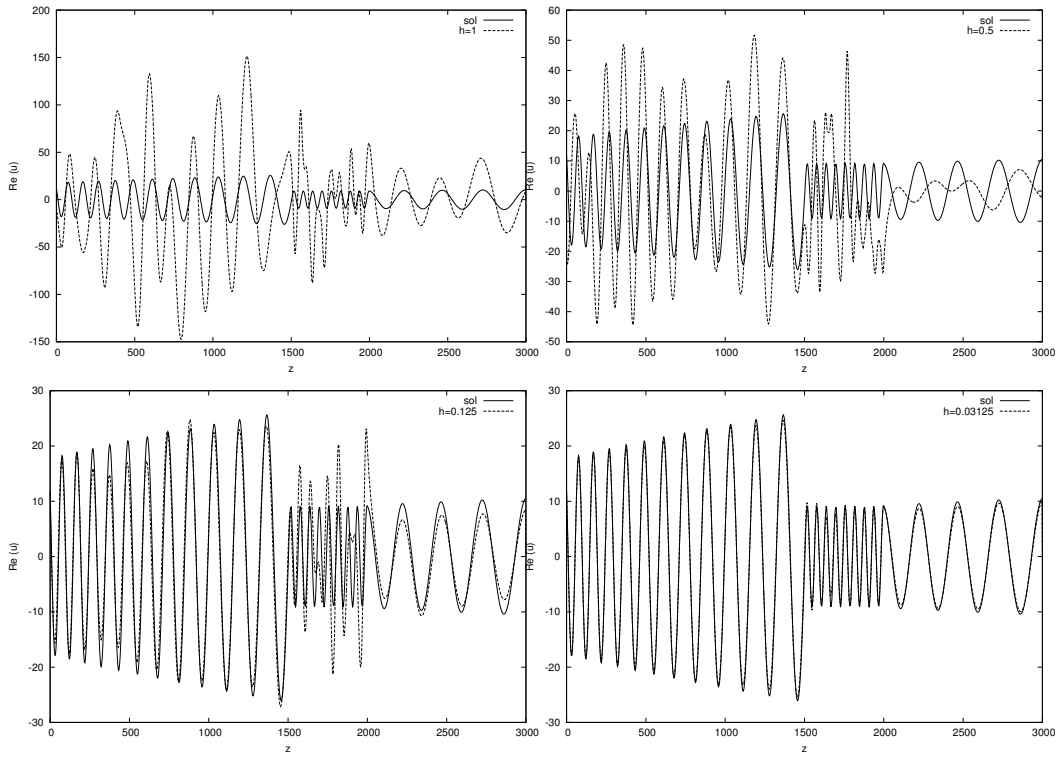


FIGURE 11. Solution profile in highly heterogeneous domain for $\mathcal{P}_6$ elements, $\omega = 40\pi$

p	err	h_z	ndf	nz
1	5.5×10^{-2}	1.5	12.0×10^5	14.4×10^6
2	4.0×10^{-2}	6	3.01×10^5	3.15×10^6
3	5.8×10^{-2}	12	1.70×10^5	2.06×10^6
4	5.9×10^{-2}	18.75	1.24×10^5	1.84×10^6
5	5.9×10^{-2}	20	1.70×10^5	3.12×10^6
6	5.5×10^{-2}	24	1.67×10^5	3.76×10^6

TABLE 5. Comparison of different p to obtain a 5% accuracy

squares. Hence, the mesh is formed by a regular of 300×1000 squares, each square being divided into two triangles.

If we use $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ elements on the fitting mesh, we obtain relative L^2 errors of 1.79×10^{-1}, 3.19×10^{-4} and 1.21×10^{-6}. We thus have that the $\mathcal{P}_1$ solution is not precise enough regarding the level of accuracy we target and the $\mathcal{P}_2$ and $\mathcal{P}_3$ solutions are very precise but in the same time very expensive to compute. For example, the computation of the $\mathcal{P}_2$ solution requires to invert a system with 1.20×10^6 degrees of freedom and 1.26×10^7 non-zero elements in the matrix.

We now focus on the size of the cells which obviously impacts the size of the corresponding linear system. It turns out that if p is greater than 2, the MMAm delivers 5% relative error on a much coarser (and non-fitting) mesh than the fitting mesh as shown in Table 5. We see that when p is greater than 2, we can use a coarse non-fitting mesh and use less than 1.20×10^6 degrees of freedom to get 5% of accuracy. We conclude that the MMAm enables to reduce the computational cost compared to the standard FEM on fitting meshes.

To give a comparison with another fitting mesh method, consider the coarsest fitting cartesian grid made of 300×1000 squares. It includes 6.01×10^5 edges, which means that lowest order DGM plane wave method would require at least 6.01×10^5 degrees of freedom to solve (see, for

example [4]). On the other hand, the $\mathcal{P}_4$ solution is computed on a 64×160 non-fitting cartesian grid. This grid is much coarser than the 300×1000 fitting grid and the number of degrees of freedom required to obtain the $\mathcal{P}_4$ solution is 1.24×10^5 (4.8 times less than for the planewave method).

Conclusion

We have proposed a robust and efficient approach to take into account fine scale variations of the velocity on a coarse mesh, the so-called Multiscale Medium Approximation method (MMAm). The numerical examples we have performed show that the MAM provides improved numerical solutions as compared to solutions based on classical FEM or parameter averaging. More precisely, the numerical example of subsection 4.4 shows that our approach gives reliable results when classical FEM or parameter averaging do not, even for high frequency.

We also have implemented a 3D MMAm solver with medium approximation which has been successfully tested on geophysical benchmarks. Future works include these performance assessments, sharper stability conditions taking into account high order polynomials, and the analysis of the continuous problem for more general velocity parameters.

References

[1] A. Abdulle, M.J. Grote, and C. Stohrer. Fe heterogeneous multiscale method for long time wave propagation. Comptes Rendus Mathématique (Académie des Sciences), vol. 351, p. 495-499, 2013.

[2] M. Ainsworth. Discrete dispersion relation for hp-version finite element approximation at high wave number. SIAM J. Numer. Anal., 42(2):553–575, 2004.

[3] M. Ainsworth, P. Monk, and W. Muniz. Dispersive and dissipative properties of discontinuous galerkin finite element methods for the second-order wave equation. J. Sci. Comput. 27 (1-3) 5-40, 2006.

[4] M. Amara, R. Djellouli, and C. Farhat. Convergence analysis of a discontinuous galerkin method with plane wave method and lagrange multipliers for the solution of helmholtz problems. SIAM J. Numer. Anal., 47:1038–1066, 2009.

[5] Y. Capdeville, L. Guillot, and J. J. Marigo. 2-d nonperiodic homogenization to upscale elastic media for p-sv waves. Geophys. J. Int, 182:903–922, 2010.

[6] Y. Capdeville, M. Zhao, and P. Cupillard. Fast fourier homogenization for elastic wave propagation in complex media. Wave Motion, 54:170–186, 2015.

[7] P.G. Ciarlet. The finite element method for elliptic problems. North-Holland, Amsterdam, 1978.

[8] P. Cummings and X. Feng. Sharp regularity coefficient estimates for complex-valued acoustic and elastic helmholtz equations, math. Models Methods Appl. Sci., 16(1):139–160, 2006.

[9] X. Feng and H. Wu. Discontinuous galerkin methods for the helmholtz equation with large wave numbers. SIAM J. Numer. Anal., 47:2872–2896, 2009.

[10] U. Hetmaniuk. Stability estimates for a class of helmholtz problems. Communications in Mathematical Sciences, 5(3):665–678, 2007.

[11] G. Hu, X. Liu, F. Qu, and B. Zhang. Variational approach to scattering by unbounded rough surfaces with neumann and genralized impedance boundary conditions. Commun. Math. Sci., 13(2):511–537.

[12] F. Ihlenburg and I. Babuška. Finite element solution of the helmholtz equation with high wave number part i: The h-version of the fem. SIAM J. Numer. Anal., 34:315–358, 1997.

[13] L.M. Imbert-Gérard and B. Després. A generalized plane wave numerical method for smooth non constant coefficients. hal-00738211v2, 2011.

[14] P. Ciarlet jr. and C. Sthorer. Finite element heterogeneous multiscale method for the classical helmholtz equation. preprint, 2014.

[15] 0. Laghrouche, P. Bettess, E. Perrey-Debain, and J. Trevelyan. Wave interpolation finite elements for helmholtz problems with jumps in the wave speed. Comput. Methods Appl. Engrg, 194:367–381, 2005.

[16] A. Lechleiter and H. Haddar. Electromagnetic wave scattering from rough penetrable layers. SIAM Journal on Mathematical Analysis, 43(5):2418–2443, 2011.

[17] A. Lechleiter and S. Ritterbusch. A variational method for wave scattering from penetrable rough layers. IMA Journal of Applied Mathematics, 75(3), 2010.

[18] T. Luostari, T. Huttunen, and P. Monk. Improvements for the ultra weak variational formulation. Int. J. Numer. Meth. Engng, 94:598–624, 2013.

[19] P. Massimi, R. Tezaur, and C. Farhat. A discontinuous enrichment method for three-dimensional multiscale harmonic wave propagation problems in multi-fluid and fluid-solid media. Int. J. Numer. Meth. Engng, 00:1–6, 2000.

[20] V. Mattesi, H. Barucq, and J. Diaz. Prise en compte de vitesses de propation polynomiales dans un code de simulation galerkine discontinue. hal-01176854v1, 2011.

[21] J.M. Melenk. On generalized finite element methods. Ph. D. thesis, University of Maryland, available at http://www.math.tuwien.ac.at/˜melenk/publications, 1995.

[22] J.M. Melenk and S. Sauter. Convergence analysis for finite element discretizations of the helmholtz equation with dirichlet-to-neumann boundary conditions. Math. Comp., 79:1871–1914, 2010.

[23] J.M. Melenk and S. Sauter. Wavenumber explicit convergence analysis for finite element discretizations of the helmholtz equation. SIAM J. Numer. Anal., 49:1210–1243, 2011.

[24] R. Tezaur, I. Kalashnikova, and C. Farhat. The discontinuous enrichment method for medium-frequency helmholtz problems with a spatially variable wavenumber. Comput. Methods Appl. Engrg, 268:126–140, 2014.

[25] H. Wu. Preasymptotic error analysis of cip-fem and fem for helmholtz equation with high wave number. part i: Linear version. IMA J. Numer. Anal., doi: 10.1093/imanum/drt033, published online: Oct. 2013.

[26] L. Zhu and H. Wu. Preasymptotic error analysis of cip-fem and fem for helmholtz equation with high wave number. part ii: *hp* version. SIAM Journal on Numerical Analysis, 51(3):1828–1852, 2013.

APPENDIX

Appendix A. Proof of lemma 1

Proof. In the proof of (4) and (5), we use the identity

$$2\mathrm{Re}\ v\partial_j\bar{v} = \partial_j|v|^2, \quad \forall v \in H^1(\Omega,\mathbb{C}),\ j=1,2. \tag{21}$$

To demonstrate (4), we first develop the expression:

$$\begin{aligned}
\partial_j w\partial_j(\mathbf{x}\cdot\nabla\bar{w}) &= \sum_{k=1}^{2}\partial_j w\partial_j(\mathbf{x}_k\partial_k\bar{w}) \\
&= \sum_{k=1}^{2}\partial_j w(\partial_j\mathbf{x}_k\partial_k\bar{w}+\mathbf{x}_k\partial_{jk}\bar{w}) \\
&= \sum_{k=1}^{2}\delta_{jk}\partial_j w\partial_k\bar{w}+\sum_{k=1}^{2}\mathbf{x}_k\partial_j w\partial_{jk}\bar{w} \\
&= |\partial_j w|^2+\sum_{k=1}^{2}\mathbf{x}_k\partial_j w\partial_k(\partial_j\bar{w})
\end{aligned}$$

Using (21) with $v=\partial_j w$, we get

$$\begin{aligned}
2\mathrm{Re}\ \partial_j w\partial_j(\mathbf{x}\cdot\nabla\bar{w}) &= 2|\partial_j w|^2+\sum_{k=1}^{2}x_k\partial_k|\partial_j w|^2 \\
&= 2|\partial_j w|^2+\mathbf{x}\cdot\nabla|\partial_j w|^2.
\end{aligned}$$

We shall now integrate and then use a Green formula:

$$\begin{aligned}
2\mathrm{Re}\int_\Omega\partial_j w\partial_j(\mathbf{x}\cdot\nabla\bar{w}) &= 2\int_\Omega|\partial_j w|^2+\int_\Omega\mathbf{x}\cdot\nabla|\partial_j w|^2 \\
&= 2\int_\Omega|\partial_j w|^2-\int_\Omega\mathrm{div}\,\mathbf{x}|\partial_j w|^2+\int_{\partial\Omega}\mathbf{x}\cdot\mathbf{n}|\partial_j w|^2 \\
&= 2\int_\Omega|\partial_j w|^2-\int_\Omega 2|\partial_j w|^2+\int_{\partial\Omega}\mathbf{x}\cdot\mathbf{n}|\partial_j w|^2 \\
&= \int_{\partial\Omega}\mathbf{x}\cdot\mathbf{n}|\partial_j w|^2
\end{aligned}$$

We demonstrate (4) by summing over j. We now turn to (5).

$$\begin{aligned}
2\mathrm{Re}\int_\Omega k^2 w\mathbf{x}\cdot\nabla\bar{w} &= \int_\Omega k^2\mathbf{x}\cdot\nabla|w|^2 \\
&= \sum_{r=1}^{R}k_r^2\int_{\Omega_r}\mathbf{x}\cdot\nabla|w|^2 \\
&= \sum_{r=1}^{R}k_r^2\left\{-\int_{\Omega_r}\mathrm{div}\,\mathbf{x}|w|^2+\int_{\partial\Omega_r}\mathbf{x}\cdot\mathbf{n}_r|w|^2\right\} \\
&= -2\int_\Omega k^2|w|^2+\sum_{r=1}^{R}k_r^2\int_{\partial\Omega_r}\mathbf{x}\cdot\mathbf{n}_r|w|^2. \\
&= -2\int_\Omega k^2|w|^2+\sum_{r,l=1}^{R}\int_{\Omega_r\cap\Omega_l}(k_r^2\mathbf{x}\cdot\mathbf{n}_r+k_l^2\mathbf{x}\cdot\mathbf{n}_l)|w|^2+\int_{\partial\Omega}k^2\mathbf{x}\cdot\mathbf{n}|w|^2
\end{aligned}$$

□

H. Barucq, INRIA Research Centre Bordeaux Sud-Ouest IPRA, University of Pau, IPRA, BP 1155, 64013 Pau, France
E-mail address: helene.barucq@inria.fr

T. Chaumont-Frelet, INRIA Research Centre Bordeaux Sud-Ouest IPRA, University of Pau, IPRA, BP 1155, 64013 Pau, France, and Normandie Université, INSA de Rouen, LMI, Av. de l'Université, 76801 St Etienne du Rouvray cedex, France
E-mail address: theophile.chaumont_frelet@insa-rouen.fr

C. Gout, Normandie Université, INSA de Rouen, LMI, Av. de l'Université, 76801 St Etienne du Rouvray cedex, France
E-mail address: christian.gout@insa-rouen.fr